

A spatiotemporal ensemble machine learning framework for generating land use / land cover time-series maps for Europe (2000 – 2019) based on LUCAS, CORINE and GLAD Landsat

Martijn Witjes (✉ martijn.witjes@opengeohub.org)

OpenGeoHub, Wageningen, The Netherlands

Leandro Parente

OpenGeoHub, Wageningen, The Netherlands

Chris J. van Diemen

OpenGeoHub, Wageningen, The Netherlands

Tomislav Hengl

OpenGeoHub, Wageningen, The Netherlands

Martin Landa

Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Lukas Brodsky

Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Lena Halounova

Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Josip Krizan

MultiOne, Zagreb, Croatia

Luka Antonic

MultiOne, Zagreb, Croatia

Codrina M Ilie

Technical University of Civil Engineering Bucharest, Bucharest, Romania

Vasile Craciunescu

National Meteorological Administration of Romania, Bucharest, Romania

Milan Kilibarda

Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia

Ognjen Antonijevic

Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade,
Belgrade, Serbia

Luka Glusica

GILAB, Belgrade, Serbia

Method Article

Keywords: landsat, spatial analysis, spatiotemporal, ensemble, machine learning, probability, uncertainty, land use, land cover, big data, environmental monitoring

DOI: <https://doi.org/10.21203/rs.3.rs-561383/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A spatiotemporal ensemble machine learning framework for generating land use / land cover time-series maps for Europe (2000 - 2019) based on LUCAS, CORINE and GLAD Landsat

Martijn Witjes^{Corresp., 1}, Leandro Parente¹, Chris J van Diemen¹, Tomislav Hengl^{1, 2}, Martin Landa³, Lukas Brodsky^{3, 4}, Lena Halounova³, Josip Križan⁵, Luka Antonić⁵, Codrina M Ilie^{6, 7}, Vasile Craciunescu^{6, 8}, Milan Kilibarda⁹, Ognjen Antonijević⁹, Luka Glušica¹⁰

¹ OpenGeoHub, Wageningen, The Netherlands

² Envirometrix, Wageningen, The Netherlands

³ Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

⁴ Mapradix, Prague, Czech Republic

⁵ MultiOne, Zagreb, Croatia

⁶ Terrasigna, Bucharest, Romania

⁷ Technical University of Civil Engineering Bucharest, Bucharest, Romania

⁸ National Meteorological Administration of Romania, Bucharest, Romania

⁹ Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia

¹⁰ GILAB, Belgrade, Serbia

Corresponding Author: Martijn Witjes
Email address: martijn.witjes@opengeohub.org

A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and analysis of land use / land cover (LULC) dynamics is presented. The framework includes: (1) harmonization and preprocessing of high-resolution spatial and spatiotemporal covariate datasets (GLAD Landsat, NPP/VIIRS) including 5 million harmonized LUCAS and CORINE Land Cover-derived training samples, (2) model building based on spatial k-fold cross-validation and hyper-parameter optimization, (3) prediction of the most probable class, class probabilities and uncertainty per pixel, (4) LULC change analysis on time-series of produced maps. The spatiotemporal ensemble model was fitted by combining random forest, gradient boosted trees, and artificial neural network, with logistic regressor as meta-learner. The results show that the most important covariates for mapping LULC in Europe are: seasonal aggregates of Landsat green and near-infrared bands, multiple Landsat-derived spectral indices, and elevation. Spatial cross-validation of the model indicates consistent performance across multiple years with 62%, 70%, and 87% accuracy when predicting 33 (level-3), 14 (level-2), and 5 classes (level-1); with artificial surface classes such as 'airports' and 'railroads' showing the lowest match with validation points. The spatiotemporal model outperforms spatial models on known-year classification by 2.7% and unknown-year

classification by 3.5%. Results of the accuracy assessment using 48,365 independent test samples shows 87% match with the validation points. Results of time-series analysis (time-series of LULC probabilities and NDVI images) suggest gradual deforestation trends in large parts of Sweden, the Alps, and Scotland. An advantage of using spatiotemporal ML is that the fitted model can be used to predict LULC in years that were not included in its training dataset, allowing generalization to past and future periods, e.g. to predict land cover for years prior to 2000 and beyond 2020. The generated land cover time-series data stack (ODSE-LULC), including the training points, is publicly available via the Open Data Science (ODS)-Europe Viewer.

¹ **A Spatiotemporal Ensemble Machine**

² **Learning Framework for Generating Land**

³ **Use / Land Cover Time-series Maps for**

⁴ **Europe (2000–2019) based on LUCAS,**

⁵ **CORINE and GLAD Landsat**

⁶ **Martijn Witjes¹, Leandro Parente¹, Chris van Diemen¹, Tomislav Hengl¹,**

⁷ **Martin Landa², Lukáš Brodský², Lena Halounová², Josip Križan³, Luka**

⁸ **Antonić³, Codrina Maria Ilie⁴, Vasile Craciunescu⁴, Milan Kilibarda⁵,**

⁹ **Ognjen Antonijević⁵, and Luka Glušica⁶**

¹⁰ **¹OpenGeoHub, Wageningen, the Netherlands**

¹¹ **²Department of Geomatics, Faculty of Civil Engineering, CTU in Prague, Czech Republic**

¹² **³MultiOne, Zagreb, Croatia**

¹³ **⁴Terrasigna, Romania**

¹⁴ **⁵Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University**

¹⁵ **of Belgrade, Belgrade, Serbia**

¹⁶ **⁶GiLAB, Belgrade, Serbia**

¹⁷ Corresponding author:

¹⁸ Martijn Witjes¹

¹⁹ Email address: martijn.witjes@opengeohub.org

²⁰ **ABSTRACT**

21 A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and
22 analysis of Land Cover / Land Use dynamics is presented. The framework includes: (1) harmonization and
23 preprocessing of high-resolution spatial and spatiotemporal covariate datasets (GLAD Landsat, NPP/VIIRS)
24 including 5 million harmonized LUCAS and CORINE Land Cover-derived training samples, (2) model
25 building based on spatial k-fold cross-validation and hyper-parameter optimization, (3) prediction of the
26 most probable class, class probabilities and uncertainty per pixel, (4) LULC change analysis on time-series
27 of produced maps. The spatiotemporal ensemble model was fitted by combining random forest, gradient
28 boosted trees, and artificial neural network, with logistic regressor as meta-learner. The results show that
29 the most important covariates for mapping LULC in Europe are: seasonal aggregates of Landsat green and
30 near-infrared bands, multiple Landsat-derived spectral indices, and elevation. Spatial cross-validation of
31 the model indicates consistent performance across multiple years with 62%, 70%, and 87% accuracy when
32 predicting 33 (level-3), 14 (level-2), and 5 classes (level-1); with artificial surface classes such as "airports"
33 and "railroads" showing the lowest match with validation points. The spatiotemporal model outperforms
34 spatial models on known-year classification by 2.7% and unknown-year classification by 3.5%. Results of
35 the accuracy assessment using 48,365 independent test samples shows 87% match with the validation
36 points. Results of time-series analysis (time-series of LULC probabilities and NDVI images) suggest gradual
37 deforestation trends in large parts of Sweden, the Alps, and Scotland. An advantage of using spatiotemporal
38 ML is that the fitted model can be used to predict LULC in years that were not included in its training
39 dataset, allowing generalization to past and future periods, e.g. to predict land cover for years prior to 2000
40 and beyond 2020. The generated land cover time-series data stack (ODSE-LULC), including the training
41 points, is publicly available via the Open Data Science (ODS)-Europe Viewer.

42 Submitted to PeerJ on 7th of May 2021

43 INTRODUCTION

44 Anthropogenic land cover change has influenced global climate since the Paleolithic (Kaplan et al., 2011)
45 and continues to be a major driver of regional (Pielke Sr et al., 2002) and global (Houghton et al., 2012)
46 climate change. Furthermore, it is the single largest cause of global biodiversity loss (Sala et al., 2000),
47 and has quantifiable consequences for the availability and quality of natural resources, water, and air
48 (Foley et al., 2005). Key applications of land cover change maps are to inform policy (Duveiller et al.,
49 2020), analyse land-based emissions (Hong et al., 2021) and/or help estimate local climate extremes (Sy
50 and Quesada, 2020). Quantifying land cover dynamics is often crucial for policy-making at regional and
51 global levels (Liu et al., 2020b; Trisurat et al., 2019; Shumba et al., 2020).

52 Land cover has been traditionally first mapped by doing visual interpretation of aerial photographs;
53 later on by automating classification of multispectral remotely sensed data / with semi-supervised or
54 fully-supervised methods (Townshend et al., 2012; Feranec et al., 2016; Liu et al., 2021). There are
55 currently multiple global (Feng and Bai, 2019; Buchhorn et al., 2020) and regional (Homer et al., 2007;
56 Batista e Silva et al., 2013; Pflugmacher et al., 2019; Malinowski et al., 2020) land cover products based
57 on using Machine Learning and offering predictions (or their refinements) at high spatial resolutions
58 for the whole of continental Europe (Table 1). The increasing number of land cover applications and
59 datasets in Europe can largely be attributed to (1) the extensive Land use and Coverage Area frame Survey

60 ([LUCAS](https://land.copernicus.eu/imagery-in-situ/lucas)) *in-situ* point data being publicly available for research (at <https://land.copernicus.eu/imagery-in-situ/lucas>), and (2) [NASA](#)'s Landsat and [ESA](#)'s Sentinel multispectral images being
61 increasingly available for spatial analysis ([Szantoi et al., 2020](#); [Liu et al., 2021](#)).

63 Not all land cover prediction algorithms and systems, however, perform equally. [Vilar et al. \(2019\)](#)
64 have done extensive evaluation of accuracy of the [Coordination of Information on the Environment](#)
65 ([CORINE](#)) Land Cover ([CLC](#)) products for period 2011–2012 using the [LUCAS](#) data and found that
66 agreement with [LUCAS](#) was slightly higher for [Climate Change Initiative — Land Cover \(CCI-LC\)](#) (59%;
67 18 classes) than for [CLC](#) (56%; 44 classes). [Gao et al. \(2020\)](#) has evaluated accuracy of the global 30 m
68 resolution products [GlobeLand30](#) with 10 classes ([Chen et al., 2015](#)), and [Global Land Cover with Fine](#)
69 [Classification System at 30 m \(GLC FCS30\)](#) with 18 classes ([Zhang et al., 2020](#)) using the [LUCAS](#) point
70 data and concluded that the [GlobeLand30-2010](#) product agrees with [LUCAS](#) points up to 89%, while
71 [GLC FCS30-2015](#) agrees up to 85%. The large difference in the agreement reported by [Vilar et al. \(2019\)](#)
72 and [Chen et al. \(2015\)](#) can be attributed to the number of classes in the two studies: the absolute accuracy
73 linearly drops with the number of classes ([Herold et al., 2008](#); [Van et al., 2019](#)), and usually the accuracy
74 results for 6–10 classes vs 40 classes can be up to 50% better.

75 Overall, the land cover mapping results in Europe match the results of [Calderón-Loor et al. \(2021\)](#)
76 who estimated 90% accuracy on 6 classes with 7 years (from 1985 to 2015) of Landsat data of Australia.
77 [Tsendsazar et al. \(2018\)](#) reports similar accuracy levels for Africa. Likewise, [Liu et al. \(2020a\)](#) reports
78 83% accuracy on 7 classes with 34 years of [Global Land Surface Satellite \(GLASS\)](#) data. USA one of
79 the first continent-scale countries to produce and distribute the National Land Cover Database at 30 m
80 (<https://www.mrlc.gov/>; years 2001, 2004, 2006, 2008, 2011, 2013, 2016, 2018) ([Homer et al., 2020](#)), which focuses on 16 classes and reports accuracy of at least 80%.

82 [Inglada et al. \(2017\)](#) report a kappa of 0.86 for mapping 17 classes of land cover for France for year
83 2014. The most-up-to-date land cover products for Europe by ([Malinowski et al., 2020](#)) report mapping
84 accuracy of 86% based on predicting 13 classes with 2017 Sentinel-2 data. The [ESA](#)'s WorldCover
85 project (<https://esa-worldcover.org/>) is another global 10 m spatial resolution land cover product
86 that aims at consistent mapping accuracy of at least 75%.

Table 1. Inventory and comparison of existing land cover data products at finer spatial resolutions (≤ 300 m) available for the continental Europe.

Product / reference	Time span	Spatial resolution	Mapping accuracy	Classification system	Uncertainty / Probability
CLC	1990, 2000, 2006, 2012, 2018	100 m (25 ha)	$\leq 85\%$	44 classes	N / N
ESA CCI-LC		300-m		22 classes	N / N
(Batista e Silva et al., 2013)	2006	100-m			
S2GLC (Malinowski et al., 2020)	2017	10 m		15 classes	
Pflugmacher et al. (2019)	2014–2016	30 m	75%	12 classes	N / N
GLC FCS30 (Zhang et al., 2020)	2015, 2020	30-m			N / N
Buchhorn et al. (2020)	2015, 2016, 2017, 2018	100 m		??	N / Y
ESA WorldCover	2020	10 m	>75%	12+	N / N
ELC10 (Venter and Sydenham, 2021)	2020	10 m	>90%	8 classes	N / N
ODSE-LULC (our product)	2000, 2001, ..., 2019	30 m		33 classes	Y / Y

87 Based on these works, it can be said that the state-of-the-art land cover mapping projects primarily
88 aim at:

- 89 (a) automating process as much as possible so that land cover maps can be produced almost on monthly
90 or even daily revisit times,
91 (b) using multi-source Earth Observation data, with especial focus on combining power of the Sentinel-1
92 and 2 data ([Venter and Sydenham, 2021](#)),
93 (c) producing higher and higher spatial resolution and thematic detail data.

94 Although the modern approaches to land cover mapping listed in Table 1 report relatively high levels
95 of accuracy, we recognize several limitations of the general approach:

- 96 • The focus of common land cover classification products is often only on hard classes (the most
97 probable class); per-pixel uncertainty in predictions is often either not reported or not derived at
98 all. Mapping accuracy is provided as a general number (average performance) for the whole area,
99 although in practice prediction accuracy often varies from class to class.
100 • Most of policy-related institutions require time-series land cover data products which are compatible
101 with legacy products such as [CLC](#) and [CCI-LC](#), while most of research focuses on producing general
102 land cover maps for recent years only. It appears that many land cover mapping missions, especially
103 the global missions, tend to sacrifice complexity of the target legend for the sake of higher nominal
104 accuracy.
105 • In the case of the land cover dynamics mapping, usually no further analysis is provided to help
106 detect and quantify trends and eventually understand main drivers of land cover change.

107 Land cover data with higher thematic resolution have shown to help improve the performance of
108 subsequent change detection ([Buyantuyev and Wu, 2007](#)), as well as the performance and level of detail
109 of modeling land cover trends ([Conway, 2009](#)) and other environmental phenomena ([Castilla et al., 2009](#);
110 [Zhou et al., 2014](#)). Increasing thematic resolution while limiting the prediction to one trained classifier,
111 however, poses several challenges: (1) training a single model on multi-year data requires extensive data
112 harmonization efforts, and (2) the exponential increase of possible change types with each additional
113 predicted class complicates the manual creation of post-classification temporal consistency rules.

114 With an increasing spatial resolution and increasing extent of EO images, the gap between historic
115 land cover maps and current 10 m resolution products is growing. This makes it difficult to identify key
116 processes of land cover change over large areas ([Veldkamp and Lambin, 2001](#); [Vilar et al., 2019](#)). Hence,
117 a balanced and consistent approach is needed that can take into account both accuracy gains due to spatial
118 resolution, and applicability for time-series analysis / change detection for longer periods of time.

119 In this paper we describe a complete seamless framework for spatiotemporal prediction, uncertainty
120 assessment and analysis of the land cover dynamics using long time-series (20+ years) in a High Per-
121 formance Computing framework. We present results of modeling and predicting land cover classes for
122 continental Europe using spatiotemporal Machine Learning at 30 m spatial resolution. We fit and use a
123 single model for the whole spacetime cube of interest. This allows us to both continue predicting land
124 cover for subsequent years, and to try to back-track land cover status even prior to the year 2000, without
125 a need to collect additional training (point) data.

126 We provide, in addition, results of internal accuracy assessment, based on 5-fold spatial cross-
127 validation with refitting (Roberts et al., 2017; Lovelace et al., 2019), a comparison of spatial vs spatiotem-
128 poral models, and also results of time-series analysis on the whole data-cube (2000–2019). We use, as
129 much as possible, a consistent methodology, which implies:

- 130 1. Using consistent training data based on consistent sampling methodology and sampling intensity
131 over the complete spacetime cube of interest (LUCAS; d'Andrimont et al. (2020));
- 132 2. Using consistent / harmonized Earth Observation images based on the Global Land Analysis and
133 Discovery (GLAD) Analysis Ready Data (ARD) Landsat product (Potapov et al., 2020), Night
134 Light images NPP/VIIRS (Román et al., 2018) and similar;
- 135 3. Providing consistent statistical analysis per every pixel of the space-time cube and per each
136 probability;

137 Our modeling framework comes at high costs however: the data we have produced is about 50–100
138 times larger in size than common land cover products with the total size of about 20 TiB (Cloud-Optimized
139 GeoTIFFs). These data is both more complex to analyze and to visualize. To deal with the data size,
140 we ran all processing in a fully automated and fully optimized High Performance Computing (HPC)
141 framework. We refer to the dataset we have produced as Open Data Science Europe — Land Use / Land
142 Cover or short ODSE-LULC.

143 In the following section we describe how we prepared data, fitted models, tested spatial vs spatiotem-
144 poral models, and fitted pixel-wise space-time regressions for NDVI and probability time-series. We then
145 report the results and discuss advantages and limitations of spatiotemporal Ensemble Machine Learning
146 (EML), and suggest what we consider could be next development directions and challenges.

147 MATERIALS AND METHODS

148 Spatiotemporal Machine Learning

149 The annual land cover product for continental Europe was generated using spatiotemporal EML. The
150 general model used to predict land cover was of the form:

$$Y(\lambda, \phi, t) = f [X(\lambda, \phi, t), X'(\lambda, \phi), \varepsilon'] \quad (1)$$

151 where λ, ϕ are the longitude and latitude, t is time of observation, m is the trend, X' are the “static”
152 covariates that are assumed constant through time (such as elevation and derived variables), and ε' is the
153 stochastic component that is added to the prediction errors. The model in Eq.(1) implies that all training
154 points are overlaid and matched either in spacetime (time-series of rasters) or space only (single list of
155 covariates).

156 In this work, for dynamically changing covariates we used harmonized and gap-filled Landsat bands
157 (Blue, Green, Red, NIR, SWIR1, SWIR2 and Thermal) derived for each of the four season (hence
158 20×4 images) and Suomi-NPP Visible Infrared Imaging Radiometer Suite (VIIRS) night light images

159 down-scaled from 500 m to 30 m using cubic-splines. For static covariates, we used continental EU
160 DTM derivatives elevation and slope in percent (Hengl et al., 2021), and the 30+ year probability of water
161 occurrence (Pekel et al., 2016).

162 As an additional space-time varying covariate we used the geometric Earth surface minimum and
163 maximum temperature, which can be defined universally anywhere on globe by using (Kilibarda et al.,
164 2014):

$$t_{min} = 24.2 \cdot \cos \phi - 15.7 \cdot (1 - \cos \theta) \cdot \sin |\phi| - 0.6 \cdot \frac{z}{100} \quad (2)$$

$$t_{max} = 37 \cdot \cos \phi - 15.4 \cdot (1 - \cos \theta) \cdot \sin |\phi| - 0.6 \cdot \frac{z}{100} \quad (3)$$

165 where θ is derived as:

$$\theta = (day - 18) \cdot \frac{2\pi}{365} + 2^{1-\text{sgn}(\phi)} \cdot \pi. \quad (4)$$

166 where *day* is the day of year, ϕ is the latitude, the number 18 represents the coldest day in the northern
167 and warmest day in the southern hemisphere, z is the elevation in meter, 0.6 is the vertical temperature
168 gradient per 100 m, and sgn denotes the signum function that extracts the sign of a real number. In R
169 syntax the function from Eqs.(2–4) can be implemented as:

```
170 temp . from . geom <- function( fi , day , a=30.419375 ,  
171 b=-15.539232 , elev=0 , t . grad=0.6 ) {  
172   f = if else( fi==0 , 1e-10 , fi )  
173   costeta = cos( (day-18)*pi/182.5 + 2^(1-sign(fi)) *pi)  
174   cosfi = cos( fi*pi/180 )  
175   A = cosfi  
176   B = (1-costeta) * abs( sin( fi*pi/180 ) )  
177   x = a*A + b*B - t . grad * elev / 100  
178   return(x)  
179 }  
180 temp . from . geom( fi=52 , day=120 ) # => 8.73
```

181 The geometric minimum and maximum temperature can be considered “*geographical covariates*”
182 because they are basically geometric transformations of latitude and day of the year. We used them
183 because we assume that they can help Machine Learning algorithms distinguish between land cover
184 classes under distant latitudes e.g. coniferous forest in Greece and Norway.

185 A detailed overview of the workflow used to fit models and produce predictions of land cover is
186 presented in Fig. 1. It was implemented in Python and R programming languages, and is publicly available,
187 under an Apache-2 license, through the eumap library. The eumap package builds upon scikit learn
188 (Pedregosa et al., 2011; Géron, 2019); with StackingClassifier as the key function used to produce

189 EML.

190 For modeling we used an ensemble of three learners:

- 191 1. Random Forest ([Breiman, 2001](#));
- 192 2. Gradient-boosted trees ([Chen and Guestrin, 2016](#));
- 193 3. Artificial Neural Network ([McCulloch and Pitts, 1943](#));

194 These were selected among initial 10 learners which we tested on sample data first. We fine-tune and
195 optimize the hyperparameters for three ML models by minimizing the log-loss metric derived from a
196 5-fold spatial cross validation ([Lovelace et al., 2019](#)), based on a 30×30 km tilling system. We then fit
197 the ensemble model (meta-learner) using the best hyperparameters attached to each learner and used the
198 logistic regression classifier ([Defazio et al., 2014](#)) as the final meta-learner.

199 Although [EML](#) model fitting and prediction is at the order of magnitude more computational than
200 using a single learner (as in [Malinowski et al. \(2020\)](#) and/or [Venter and Sydenham \(2021\)](#)), it has shown
201 to bring several advantages: (1) it typically helps increase accuracy ([Seni and Elder, 2010; Zhang and Ma,](#)
202 [2012](#)), (2) it can be used to get a more reliable model-free estimate of the prediction uncertainty.

203 After the model fine-tuning and feature selection using scikit learn, we generated a single ensemble
204 model that was then used to predict for every pixel for each of the 20 years:

- 205 1. Probability per class (33 land cover classes);
- 206 2. the most probable land cover class;
- 207 3. probability error (uncertainty) per class derived as the standard deviation of three predicted proba-
208 bilities for each pixel.

209 Based on the predictions of the land cover classes, probabilities and input Landsat images, we also
210 derive for each pixel:

- 211 1. Land cover change class per year and for the 2000–2019 range;
- 212 2. slope of change fitted using logistic regression on the probabilities for key classes;
- 213 3. slope of change fitted using logistic regression for [NDVI](#) seasonal time-series.

214 All the output predictions (20 dominant land cover class, 660 per-class probabilities, and 660 per-class
215 uncertainties) were predicted first per tile, then exported as Cloud Optimized Geotiffs (COGs) files
216 and are publicly available through the Open Data Science Europe (ODS-Europe) Viewer, the S3 Cloud
217 Object Service, and from <http://doi.org/10.5281/zenodo.4725429>. The classification matrix
218 with all training points and values of covariates is available from [http://doi.org/10.5281/zenodo.](http://doi.org/10.5281/zenodo.4740691)
219 [4740691](#).

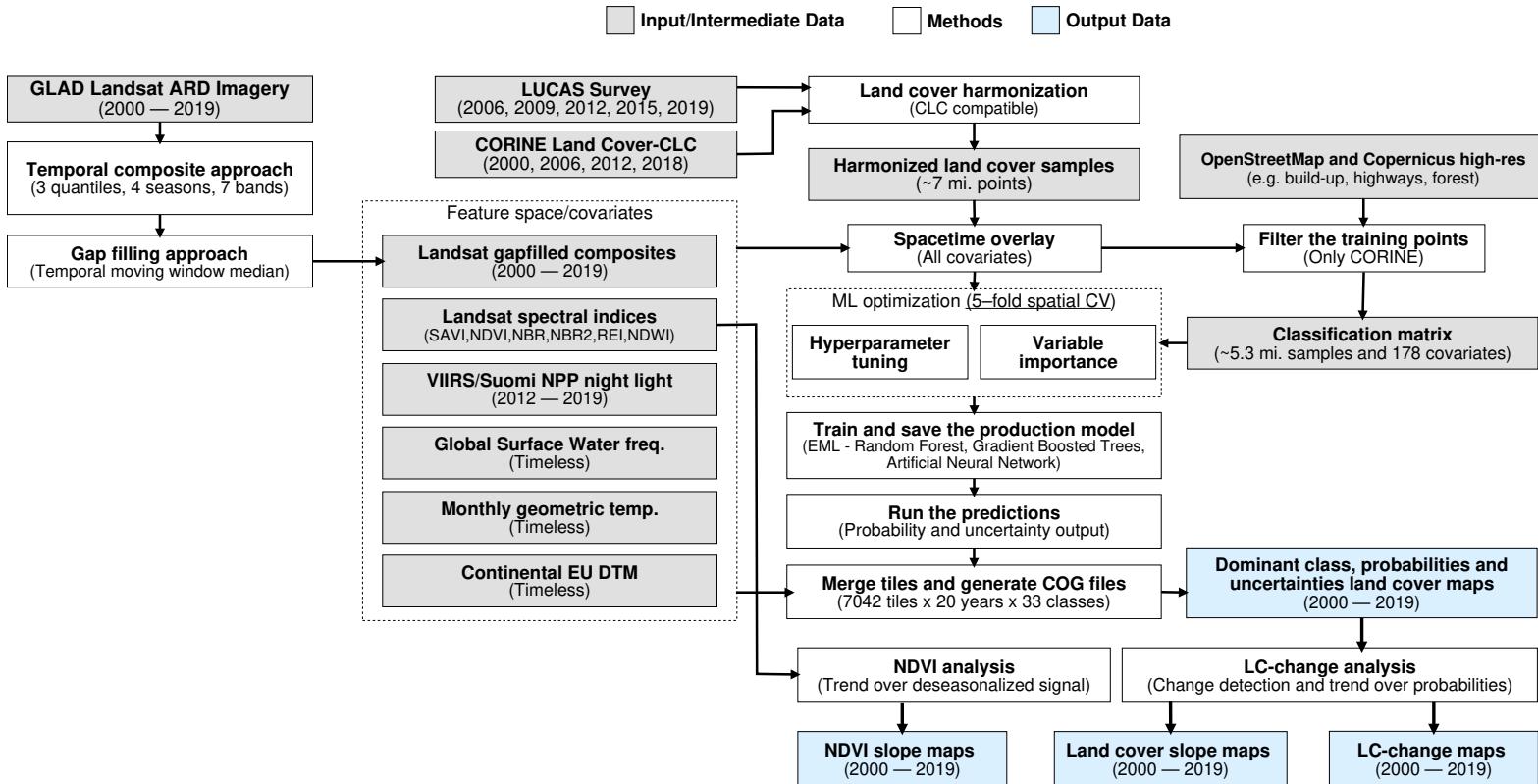


Figure 1. General workflow used to prepare point data and covariate layers, fit models and generate annual land cover products (2000–2019). Components of the workflows are described in detail via the GitLab repository of the GeoHarmonizer project (https://gitlab.com/geoharmonizer_inea/).

220 **Target land cover classification system**

221 The target land cover nomenclature was designed based on [CLC](#) nomenclature ([Bossard et al., 2000](#))
222 and is available in Table 2. [CLC](#) is probably the most comprehensive and detailed European land cover
223 product to date. The [CLC](#) program was established in 1985 by the [European Commission \(EC\)](#) to provide
224 geographically harmonized information concerning the environment on the continent. The original [CLC](#)
225 dataset is mapped in 44 classes with a minimum mapping unit of 25 ha for areal phenomena and 10 ha for
226 changes. [CLC](#) mapping relies on harmonized protocol and guidelines that are shared for country-wise
227 visual photo-interpretation.

228 Note that [CLC](#) nomenclature is complex with many classes being both land use and land cover and
229 having nested or mixed relationships. For example, “*airports*” is a land use category that can include
230 grasslands and roads. Similar confusions may appear between class pairs such as construction sites with
231 bare rock; beaches, dunes, sands with sparsely vegetated areas; vineyards with fruit trees and olive trees;
232 pastures with natural grasslands.

233 The [ODSE-LULC](#) nomenclature is based on the [CLC](#) legend, although a number of changes was
234 applied. First, we removed heterogeneous and mixed classes defined for polygon mapping (sport and
235 leisure facilities, complex cultivation patterns, land principally occupied by agriculture, agro-forestry
236 areas, etc.). Next, we marked a set of [CLC](#) classes, which are potentially problematic for pixel-wise
237 classification models and may require a contextual classification approach, for instance: Road and rail
238 network, port areas, airports, water courses and water bodies, inland wetlands and maritime wetlands,
239 coastal lagoons and estuaries. We did not remove these classes beforehand to avoid overly reducing our
240 thematic resolution, and to avoid our expectations to introduce bias in our approach. The final legend in
241 Table 2 can thus be considered a necessary compromise between being as unbiased as possible towards
242 the capabilities of our approach, maintaining compatibility to the [CLC](#) legend, and limiting computational
243 costs.

244 **Covariates**

245 All covariates used by our model are derived from remotely sensed earth observation data from multiple
246 sources, the largest share being derived from Landsat imagery. This was obtained by downloading the
247 Landsat [ARD](#), provided by [GLAD](#) ([Potapov et al., 2020](#)), for the years 1999 to 2020 and for the entire
248 extent of continental Europe (see eumap landmask ([Hengl et al., 2021](#))). This imagery archive was
249 screened to remove the cloud and cloud shadow pixels, maintaining only the quality assessment-QA
250 values labeled as clear-sky according to [GLAD](#). Second, we aggregated the individual images by season
251 according to three different quantiles (25th, 50th and 75th) and the following calendar dates for all period:

- 252 • Winter: December 2 of previous year until March 20 of current year,
- 253 • Spring: March 21 until June 24 of current year,
- 254 • Summer: June 25 until September 12 of current year,
- 255 • Fall: September 13 until December 1 of current year,

Table 2. The ODSE-LULC land cover legend used based on CLC (Bossard et al., 2000), and number of LUCAS / CLC training samples. The distribution of training samples is shown in Fig. 3.

Class name	Class description	Number of samples		
		LUCAS	CLC	Total
111: Urban fabric	The aggregated continuous and discontinuous urban fabric class dominated by urban structures, where impermeable features cover 30–100% of the land.	35.613	701.402	737.015
122: Road and rail networks	Motorways and railways, including associated installations.	27.892	13.783	41.675
123: Port areas	Infrastructure of port areas, including quays, dockyards and marinas.	-	4.265	4.265
124: Airports	Airports installations: runways, buildings and associated land.	109	6.682	6.791
131: Mineral extraction sites	Areas of open-pit extraction of construction materials (sandpits, quarries) or other minerals (open-cast mines).	14.969	40.011	54.980
132: Dump sites	Public, industrial or mine dump sites.	29	6.207	6.236
133: Construction sites	Spaces under construction development, soil or bedrock excavations, earthworks.	532	6.954	7.486
141: Urban green	Areas with vegetation within urban fabric.	22.269	17.804	40.073
211: Non-irrigated arable land	Cultivated land parcels under rain-fed agricultural use for annually harvested non-permanent crops, normally under a crop rotation system.	318.971	755.454	1.074.425
212: Permanently irrigated arable land	Cultivated land parcels under agricultural use for arable crops that are permanently or periodically irrigated.	-	44.770	44.770
213: Rice fields	Cultivated land parcels prepared for rice production, consisting of periodically flooded flat surfaces with irrigation channels.	1.773	3.734	5.507
221: Vineyards	Areas planted with vines.	15.080	69.215	84.295
222: Fruit trees and berry plantations	Cultivated parcels planted with fruit trees and shrubs, including nuts, intended for fruit production.	18.566	63.667	82.233
223: Olive groves	Cultivated areas planted with olive trees, including mixed occurrence of vines on the same parcel.	19.381	51.402	70.783
231: Pastures	Meadows with dispersed trees and shrubs occupying up to 50% of surface characterized by rich floristic composition.	204.042	704.306	908.348
311: Broad-leaved forest	Vegetation formation composed principally of trees, including shrub and bush understorey, where broad-leaved species predominate.	185.954	751.158	937.112
312: Coniferous forest	Vegetation formation composed principally of trees, including shrub and bush understorey, where coniferous species predominate.	147.552	739.288	886.840
321: Natural grasslands	Grasslands under no or moderate human influence. Low productivity grasslands. Often in areas of rough, uneven ground, also with rocky areas, or patches of other (semi-)natural vegetation.	56.440	293.825	350.265
322: Moors and heathland	Vegetation with low and closed cover, dominated by bushes, shrubs (heather, briars, broom, gorse, laburnum etc.) and herbaceous plants, forming a climax stage of development.	114.228	178.600	292.828
323: Sclerophyllous vegetation	Bushy sclerophyllous vegetation in a climax stage of development, including maquis, matorral and garrigue.	-	38.683	138.683
324: Transitional woodland-shrub	Transitional bushy and herbaceous vegetation with occasional scattered trees. Can represent either woodland degradation or forest regeneration / re-colonization.	-	724.874	724.874
331: Beaches, dunes, sands	Natural un-vegetated expanses of sand or pebble/gravel, in coastal or continental locations, like beaches, dunes, gravel pads.	9.900	15.275	25.175
332: Bare rocks	Scree, cliffs, rock outcrops, including areas of active erosion.	8.138	67.173	75.311
333: Sparsely vegetated areas	Areas with sparse vegetation, covering 10–50% of the surface.	-	221.421	221.421
334: Burnt areas	Areas affected by recent fires.	-	2.175	2.175
335: Glaciers and perpetual snow	Land covered by ice or permanent snowfields.	319	6.954	7.273
411: Inland wetlands	Low-lying land usually flooded in winter, partly saturated by water; and wetlands with considerable amount of decomposed moss and vegetation matter.	17.158	256.401	273.559
421: Maritime wetlands	Vegetated low-lying coastal areas above the high-tide line, susceptible to seawater flooding; salt-pans for salt extraction, and tidal coastal zones.	1.094	15.126	16.220
511: Water courses	Natural or artificial water courses for water drainage channels.	6.030	8.582	14.612
512: Water bodies	Natural or artificial water surfaces covered by standing water most of the year.	20.038	198.979	219.017
521: Coastal lagoons	Stretches of salt or brackish water in coastal areas which are separated from the sea by a tongue of land or other similar topography.	1.401	2.270	3.671
522: Estuaries	The mouth of a river under tidal influence within which the tide ebbs and flows.	-	1.241	1.241
523: Sea and ocean	Zone seaward of the lowest tide limit.	-	595	595

256 From more than 73 TiB of input data we produced 84 images (3 quantiles \times 4 seasons \times 7 Landsat
257 bands) for each year with different occurrences of no-data values due to cloud contamination in all
258 observations of a specific season.

259 We next impute all missing values in the Landsat temporal composites using the “*Temporal Moving*
260 *Window Median*” **TMWM** algorithm we implemented in python. The algorithm works as follows: we
261 first detect gaps and artifacts in the Landsat data, next we iteratively interpolate values from the temporal
262 neighbours until all pixels have values using the median function. For gap-filling we prioritize observations
263 of: 1—the same season, 2—neighboring seasons and 3—all the year.

264 The **TMWM** has been selected among 3–4 alternative algorithms by bench-marking, and to our
265 knowledge provides the best combination of gap-filling accuracy and computational costs. It is publicly
266 available in the eumap library.

267 In addition to the Landsat data, we also used as covariates:

- 268 • Landsat spectral indices: [Normalized Difference Vegetation Index \(NDVI\)](#), [Soil Adjusted Vegetation](#)
269 [Index \(SAVI\)](#), [Modified Soil Adjusted Vegetation Index \(MSAVI\)](#), [Normalized Difference Moisture](#)
270 [Index \(NDMI\)](#), Landsat [Normalized Burn Ratio \(NBR\)](#), [NBR2](#), and [Normalized Difference Water](#)
271 [Index \(NDWI\)](#) derived according to formulas in Table 3;
- 272 • Time-series SUOMI NPP [VIIRS](#) night lights at 500 m resolution downsampled to 30 m resolution
273 ([Hillger et al., 2013](#));
- 274 • Global surface water frequency at 30 m resolution ([Pekel et al., 2016](#));
- 275 • Continental EU [DTM](#)-based elevation and slope ([Hengl et al., 2021](#));
- 276 • Geometric minimum and maximum temperature derived for every pixel using Eq.(2);

277 The Landsat spectral indices and the geometric minimum and maximum temperatures were calculated
278 on-the-fly during the prediction step, avoiding the overhead to maintain this large amount of data in
279 storage media.

Table 3. Spectral indices derived from the Landsat data and used as additional covariates in the
280 spatiotemporal EML.

Spectral Index	Equation	Reference
NDVI	$\frac{nir - red}{nir + red}$	(Tucker, 1979)
SAVI	$\frac{nir - red}{(nir + red + 0.5) \times 1.5}$	(Huete, 1988)
MSAVI	$\frac{(2 \times nir + 1) - \sqrt{(2 \times nir + 1)^2 - 8 \times (nir - red)}}{2}$	(Qi et al., 1994)
NDWI	$\frac{green - swir2}{green + swir2}$	(Gao, 1996)
NBR	$\frac{nir - thermal}{nir + thermal}$	(Key and Benson, 1999)
NDMI	$\frac{nir - swir1}{nir + swir1}$	(Jin and Sader, 2005)
NBR2	$\frac{swir1 - thermal}{swir1 + thermal}$	(Key and Benson, 2006)

282 Training points

283 Point data preparation

284 We obtained the training dataset from the geographic location of **LUCAS** (in-situ source) and the centroid
285 of all **CLC** polygons (as shown in Fig. 2), harmonized according to the 33 land cover classes (see Table 2)
286 and organized by year, where each unique combination of longitude, latitude and year was considered as
287 a independent sample, resulting in more than 7 million training points.

288 The **LUCAS** data from 2006, 2009, 2012, 2015 and 2018, as provided by Eurostat (obtained from:
289 <https://ec.europa.eu/eurostat/web/lucas>) is the largest and most comprehensive in-situ land
290 cover data set for Europe. The survey has evolved since 2000 and requires harmonisation before it can be
291 used for mapping over several years. We imported data sets from individual years and harmonized these
292 before merging it into one common database with an automated workflow implemented in Python and
293 SQL (Fig. 1). For the multi-year harmonization procedure we first harmonized attribute names, re-coded
294 variables, harmonized point locations, and aggregated the points based on their location in space and
295 time. After these operations, we translated the **LUCAS** land cover nomenclature to the **ODSE-LULC**
296 nomenclature, Table 2, according to the method designed by **Buck et al. (2015)**.

297 We assigned the **LUCAS** points with a unique land-cover class a confidence rating of 100%, while
298 the **CLC** points received 85% confidence. These confidence weights were considered in the ML training
299 step and the points were used in the spatiotemporal overlay. For example, a sample from 2018 would retrieve
300 Landsat composites with reference to the same year. The distribution of all training points is shown in
301 Fig. 3.

302 Filtering

303 The **CLC** minimal mapping unit of 25 ha required filtering on the training points before they could be
304 used to represent 30 m resolution **LULC**, for example, to remove points for “111: urban fabric” located

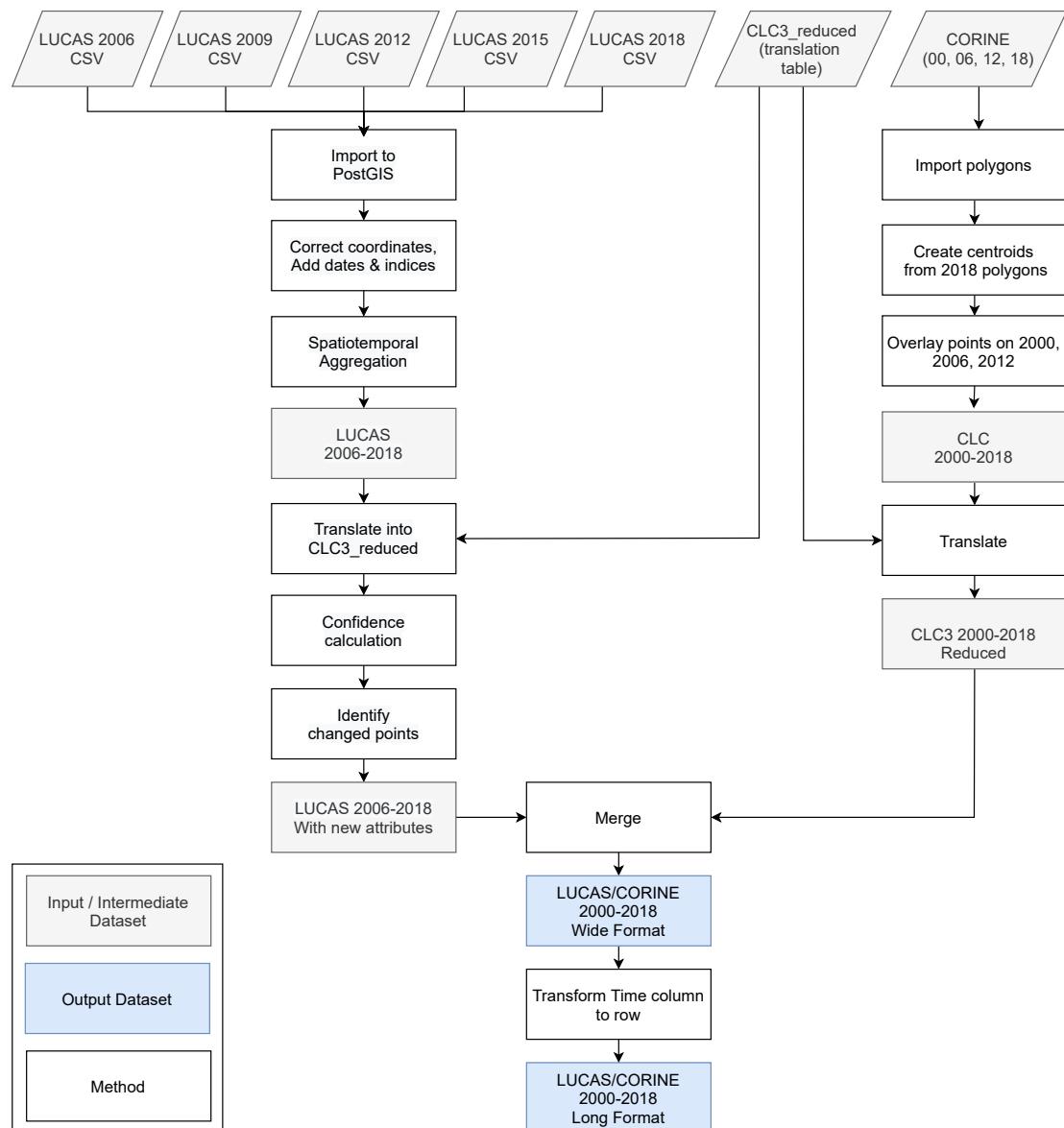


Figure 2. General workflow for merging training points obtained from LUCAS and CLC.

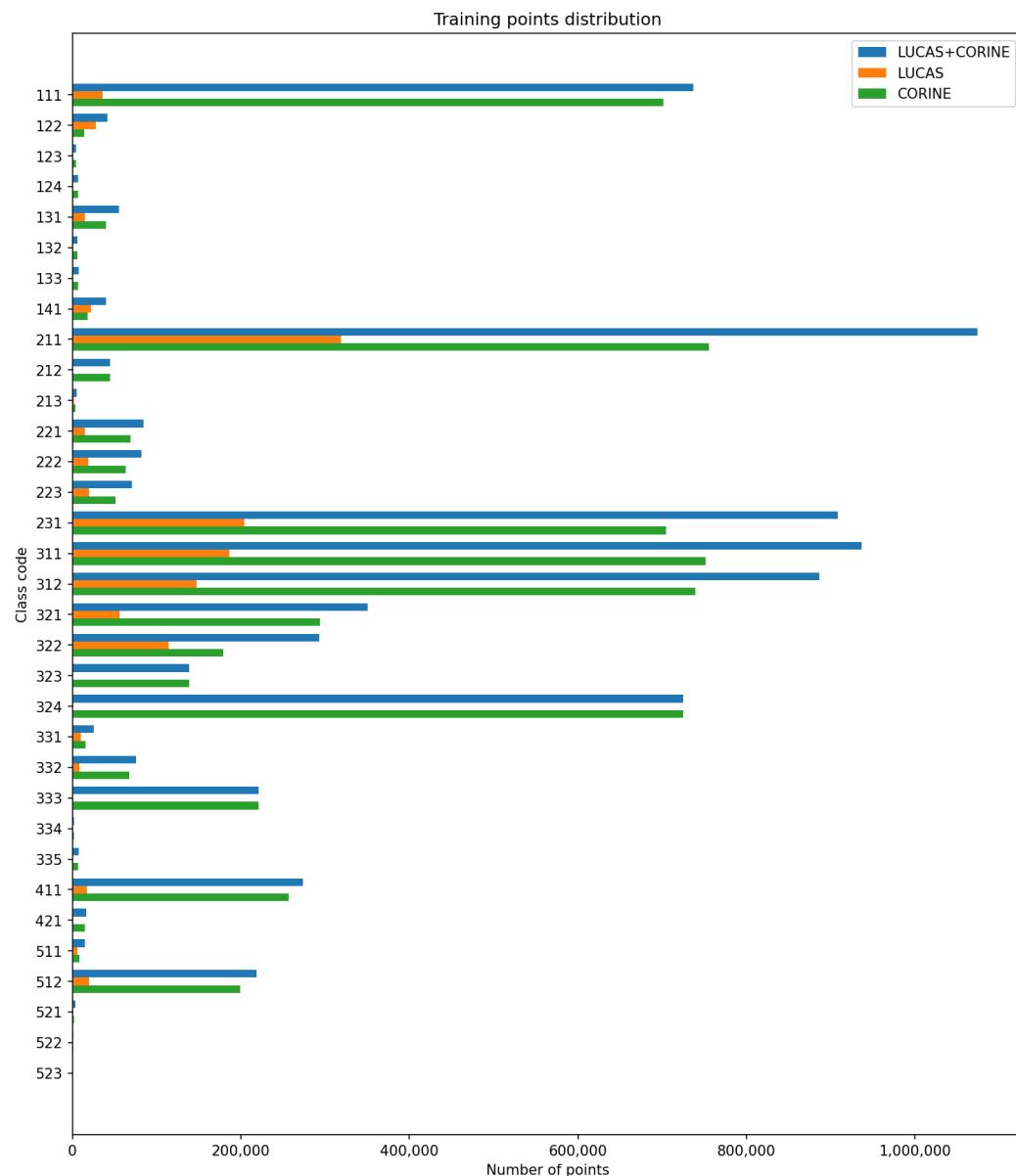


Figure 3. Training points distribution based on LUCAS, CLC and their combination. Class codes are described in Table 2.

305 in small patches of urban greenery (<25 ha). For this purpose, we extracted vector data from **OSM** layers
 306 for roads, railways, and buildings (obtained from <https://download.geofabrik.de/>), filtering them
 307 according to their crowd source-derived names and the following steps:

- 308 • Extract all possible values of the most descriptive variable in the **OSM** data;
- 309 • Sum the number of occurrences and the total area covered by vector features of each type over all
 310 countries;
- 311 • Assign each type to a category until at least 99% of vector features are categorized.

312 After this categorization step, the **OSM** layers were first rasterized to 10 m resolution by assigning
 313 cells located in vector features the value 100. These rasters were then averaged to 30 m to create a 0—100
 314 density layer for the three feature types. The building density raster was then combined with Coperni-
 315 cus **High Resolution Layers (HRL)** (obtained from <https://land.copernicus.eu/pan-european/high-resolution-layers>) to supplement its coverage where crowd-sourced **OSM** information was
 317 unavailable.

318 In the **OSM**, buildings are digitized at highest level of detail but unfortunately is incomplete for
 319 whole of Europe. To improve the coverage of the built-up areas, we first fill the non-mapped areas in
 320 **OSM** with the Impervious Built-up 2018 pixel values, which was averaged to 30 m and rescaled to
 321 101—200 values. We also use **HRL** products to filter other classes: Table 4 shows the exact conditions
 322 points of specific **LULC** classes needed to meet in order to be retained in our dataset. This is a similar
 323 procedure to the one used by [Inglada et al. \(2017\)](#). This filtering process removed about 1.3 million
 324 points from our training dataset, resulting in a classification matrix with a total of ca. 5.3 million
 325 samples and 178 covariates. The classification matrix used to produce **ODSE-LULC** is available from
 326 <http://doi.org/10.5281/zenodo.4740691>.

Table 4. Per-class conditions applied only to **CLC** points during the filtering step. All the raster layers were upsampled to 30 m² by average and the points that did not meet the specified condition were omitted from the training dataset.

Class	Condition	Copernicus HRL						OSM		HRL + OSM	
		Tree Cover	Grasslands	Imperviousness	Permanent Water	Permanent Wetness	Temporary Wetness	Railways	Roads	Buildings	
111	-										>50 and <150
122	OR			>30				>30	>30		
131	AND	equals 0	equals 0								
141	OR	>0	>0								
211	AND	equals 0									
221	AND		equals 0								
222	AND		equals 0								
223	AND		equals 0								
231	AND	equals 0									
311	AND	>0									
312	AND	>0									
321	AND	equals 0									
411	OR				>0		>0				
512	-			equals 100							

327 Comparison with other land cover products

328 We compare a number of existing land cover products to the harmonized training dataset used by our
 329 model to (see Table 1):

- 330 • GLC FCS30–2015 (Zhang et al., 2020);
- 331 • GLC FCS30–2020 (Zhang et al., 2020);
- 332 • S2GLC (Malinowski et al., 2020);
- 333 • The European land cover product for 2015 created by Pflugmacher et al. (2019);
- 334 • ELC10 (Venter and Sydenham, 2021).

335 For each comparison, we reclassified the training dataset to the nomenclature of the target dataset. We
336 overlaid our LUCAS and CLC training points from years adjacent to each land cover product. We then
337 performed a classification accuracy assessment, calculating the associated accuracy metrics. Points with
338 classes of the target products that were completely absent in the training point subsets (due to the target
339 nomenclature of the training points) were removed before these assessments.

340 The GLC FCS30 nomenclature was not suitable for direct translation because some land cover
341 groups (such as forests) are separated in several subcategories. We therefore aggregated their thematic
342 resolution to the higher level of abstraction described in Zhang et al. (2020). The complete translation
343 scheme is available via the GitLab repository of the GeoHarmonizer project (https://gitlab.com/geoharmonizer_inea/spatial-layers).

345 Accuracy assessment

346 We assess performance of our final ensemble model in two ways. Firstly through spatial 5-fold cross-
347 validation, and secondly by validating the final model predictions on an independently collected test
348 dataset. In all comparisons and experiments, we discriminate model performance with the Weighted
349 F1-score metric (Van Rijsbergen, 1980):

$$\text{WF}_1 = \sum_{c=1}^n S_c \cdot \frac{2 \cdot P_c \cdot R_c}{P_c + R_c} \quad (5)$$

350 where n is the number of classes, and S_c is the support, P_c the precision, and R_c the recall of a given class
351 c . We used this metric because it distinguishes classification performance more strictly on imbalanced
352 datasets.

353 Spatial cross-validation

354 We performed spatial 5-fold cross-validation using the hyperparameters of the final EML model and
355 assessed its performance based on its cross-validation predictions. During the hyperparameter optimization
356 step, we saved the spatial 5-fold cross-validation predictions of the model with the hyperparameters
357 that would be used to train the final model on all available data. We merged these predictions into one
358 dataset, which we treated as an independent prediction. Because no test dataset was available that had
359 the same thematic resolution as our model, we used this validation step to assess its performance on a
360 class-by-class basis. This validation strategy also allowed us to assess average model performance per
361 year. The hierarchical nature of the target nomenclature allows for accuracy assessment at three levels.

362 For each dataset size ranging from 1000 to 500,000 points, we created a training dataset from **CLC**
363 data for 2000, 2006, and 2012. Each training dataset was accompanied by a validation dataset half its
364 size with points from the same year. Lastly, for each dataset size, we created a validation set with points
365 from 2018. For each dataset size, we trained a spatial model on data from each separate year, as well as a
366 spatiotemporal model on all data from the three years.

367 **Independent test data**

368 After training an ensemble model with the same hyperparameters on all training data, we classified **LULC**
369 in 2017. This prediction was validated with the **S2GLC** dataset (Malinowski et al., 2020). This dataset
370 contains 51,926 points with human-verified land cover classifications, which were collected and used for
371 the same purpose by Malinowski et al. (2020). We performed a spatial overlay with these points on the
372 **ODSE-LULC** 2017 predictions.

373 As the **S2GLC** points follow a different nomenclature, we translated the **ODSE-LULC** predicted
374 classes according to Table 5. As our model was not trained to predict peat bogs, we removed all points
375 with this land cover class from the validation dataset. In addition, because any predicted classes outside
376 the **S2GLC** nomenclature would be automatically counted as errors, we performed two validations: (1)
377 a conservative assessment that included points with such predictions, and (2) an optimistic assessment
378 where they were omitted.

379 **Comparison of spatial and spatiotemporal models**

380 We include two experiments to compare model performance depending on the temporal nature of the
381 training data and either the data source or the size of the training dataset. In both cases we trained spatial
382 (trained on spatial data from one year) models and spatiotemporal (trained on spatial data from multiple
383 years) models, and compared their performance both when predicting on years that were included in their
384 training dataset, and on data from 2018, which we omitted from all training datasets. We averaged the
385 performance of all spatial models to obtain the performance of one '*spatial model*' approach.

386 To investigate the effect of different temporal extents and data source combinations, we compiled
387 three datasets: One with only **CLC** points, one with only **LUCAS** points, and one with points from both
388 datasets, resembling the composition of our full dataset.

389 We compared the performance of multiple spatial models trained on data from one year with the
390 performance of a spatiotemporal model trained on a separate, but equally-sized multi-year dataset. We
391 also trained a spatiotemporal model on all spatial model training data to quantify the performance gain
392 from having access to more training points than spatial models. Each model was validated on data from
393 each year, a multi-year validation set, and data from a year that was completely excluded from the training
394 datasets.

395 To investigate the effect of different temporal extents and training dataset size, we created 9 training
396 datasets from the best-performing data source with sizes ranging between 1000 and 500,000 points from
397 2000, 2006 and 2012. We accompanied each training dataset by a validation set half its size. We also
398 made nine validation datasets with data from 2018 to compare model performance on data from previously
399 unseen years. For each dataset size, three spatial models were trained on data from the three separate
400 years, and a spatiotemporal model was trained on all data for each year. The models were validated on the

Table 5. Reclassification key used to validate the predictions of our ensemble model on the test set collected by [Malinowski et al. \(2020\)](#).

S2GLC	ODSE-LULC
Artificial surfaces	111: Urban fabric 122: Road and rail networks and associated land 123: Port areas 124: Airports 132: Dump sites 133: Construction sites
Broadleaf tree cover	311: Broad-leaved forest
Coniferous tree cover	312: Coniferous forest
Cultivated areas	211: Non-irrigated arable land 212: Permanently irrigated arable land 213: Rice fields
Herbaceous vegetation	231: Pastures 321: Natural grasslands
Marshes	411: Inland wetlands 421: Maritime wetlands
Moors and heathland	322: Moors and heathland
Natural material surfaces	131: Mineral extraction sites 331: Beaches, dunes, sands 332: Bare rocks
None (not in nomenclature)	141: Green urban areas 222: Fruit trees and berry plantations 223: Olive groves 324: Transitional woodland-shrub 333: Sparsely vegetated areas 334: Burnt areas
Peatbogs	None (not in nomenclature)
Permanent snow	335: Glaciers and perpetual snow
Sclerophyllous vegetation	323: Sclerophyllous vegetation
Vineyards	221: Vineyards
Water bodies	511: Water courses 512: Water bodies 521: Coastal lagoons 522: Estuaries 523: Sea and ocean

⁴⁰¹ same-year validation sets, as well as the unknown-year validation set.

⁴⁰² **Time-series analysis**

⁴⁰³ We analysed changes using **NDVI** and probability trend analysis, **LULC** change as hard classes, and
⁴⁰⁴ finally methods to aggregate changes over larger areas. The time-series analysis has resulted in 13 data
⁴⁰⁵ layers for **NDVI** slopes, 3 for probability slopes two layers for every two years for land use change classes
⁴⁰⁶ plus two layers for changes between the years 2001–2018 and one layer for aggregated change classes

407 and one for intensity of change. Here we discuss a subsection of examples.

408 **NDVI and LULC probability trend analysis**

409 For both the **NDVI** trend analysis and the probability trend analysis we analysed the trend over the years
410 between 2000 and 2019. **NDVI** is widely used as a proxy for chlorophyll concentration and hence as an
411 indicator for vegetation health.

412 We fitted a regression for the time-series of every pixel (e.g. 20×4 of **NDVI** images) to map the slope
413 of change through time. Before applying the **OLS** estimate of the beta coefficients, we applied a logit
414 transformation to the input data because probabilities only have meaningful values between 0 and 1 and
415 **NDVI** are only meaningful for values between -1 and 1.

416 We applied **OLS** in parallel and then saved the slope, intercept, and R-squared for each pixel in Europe.
417 For **NDVI** values we also deseasonalize the data before deriving the slope by applying a loess regression
418 for seasonal decomposition (Cleveland et al., 1990) as implemented in the python statsmodels library
419 (Seabold and Perktold, 2010). An example in Python code can be found below:

```
420 from statsmodels.tsa.seasonal import STL
421 import statsmodels.formula.api as smf
422 import pandas as pd
423 import numpy as np
424 import math

425
426 def logit(p):
427     return math.log(p/(1-p))

428
429 def slope_analysis(ndvi_timeseries):

430
431     # apply deseasonalization
432     ndvi = pd.Series(np.array(ndvi_timeseries))
433     stl = STL(ndvi, period = 4, robust = True)
434     res = stl.fit()

435
436     # apply logit transformation
437     data_norm = [logit(x) for x in res.trend]
438     df_ols = pd.DataFrame(data_norm).reset_index()
439     df_ols.columns = ['t', 's']

440
441     # apply OLS
442     model = smf.ols('s~t', df_ols)
443     results = model.fit()
444     a, b = results.params
```

446 **return** res , a , b

447 For the analysis of probability slopes a similar function is applied with only the deseasonalization
 448 removed. Fig. 4A, and B show an example of the application and result of this method on a single pixel
 449 time-series in Björnrike, Sweden. Note computation of slopes is computationally very intensive often
 450 taking couple of days of continuous computing.

451 We applied a trend analysis on probability slopes for the four most prevalent LU classes: (1) coniferous
 452 forest, (2) non-irrigated arable land, (3) broad leaved forest, and (4) pastures. Fig. 4C illustrates this
 453 process for a single pixel time-series in Björnrike, Sweden.

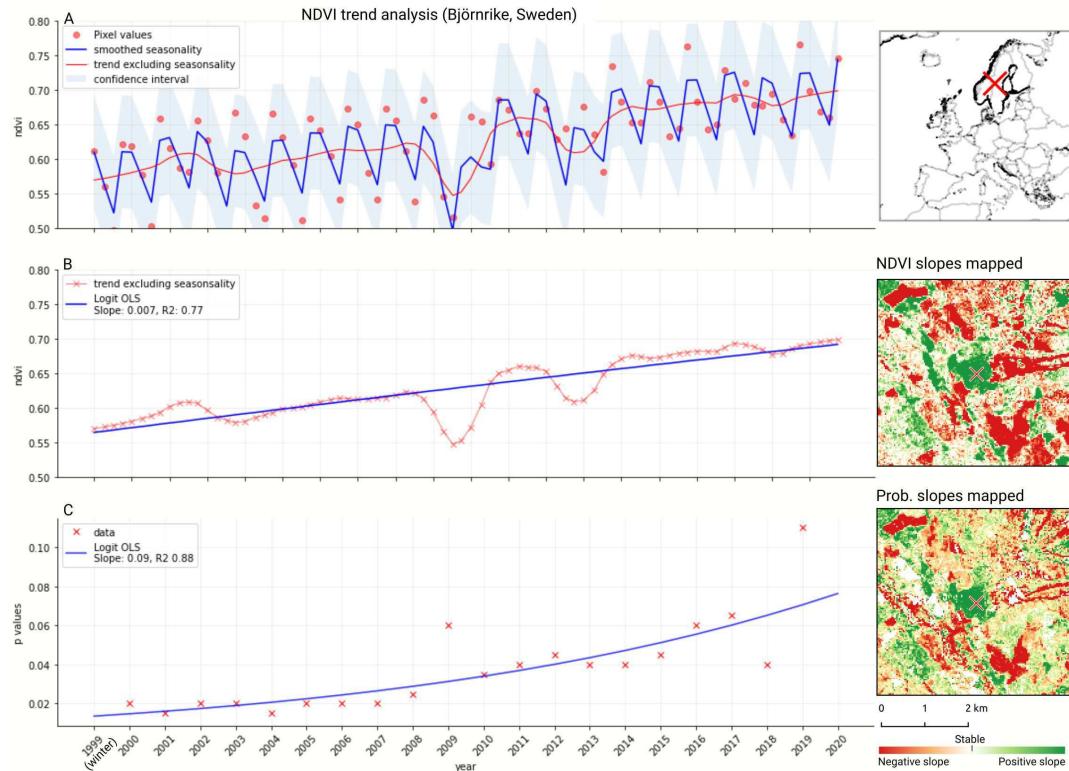


Figure 4. Example of deseasonalization (Seabold and Perktold, 2010) and subsequent Logit Ordinary Least Squares (OLS) applied on a single pixel in Sweden (coordinates: 62°24'43.7"N 13°56'00.3"E). a) red dots represent pixel values, blue line represents a local weighted regression smoothed line based on the pixel values plus a light blue area indicating the confidence interval, the red line represents the trend after removing the seasonal signal. b) red line and crosses represent the trend after removing the seasonal signal, the blue line visualizes the regression model based NDVI values in the logit space. c) Trend analysis on probability values for non-irrigated arable land. In the case above the gradient value is 0.09 with the model R-square = 0.88

454 **LULC change classes**

455 The simplest way to detect changes is to analyze difference land cover data from two years. However,
 456 this method propagates classification errors to consequent change maps and metrics (Carmel et al., 2001),
 457 which can lead to significant overestimation of change (Olofsson et al., 2014). To avoid this we follow
 458 these steps:

459 • Perform change detection on the input data before classifying land cover (Zhu and Woodcock,
460 2014);

461 • implement post-processing rules that prohibit specific types of transformations (Song et al., 2016);

462 • perform change detection and smooth over any undetected changes in predicted land cover (Li
463 et al., 2018).

464 For the hard class change analysis we applied a comparison between two subsequent years. We
465 applied one temporal post-processing step in which we consider the classification of two neighboring
466 years for every year between 2001 and 2018. In order to maximize the re-usability of our results, we
467 only performed limited post-processing: If both neighboring years are the same class and the class under
468 consideration is another class we assume this classification is an error and match the middle year with its
469 neighboring years. We call this a “*T-3 temporal filter*”. Both the filtered and the noise layers are saved for
470 further analysis and interpretation.

471 We mirror the change classes seen in the Copernicus land cover map (Buchhorn et al., 2020) and
472 described in Table 6. The change classes applied by the Copernicus land cover map, however, use classes
473 of a higher abstraction level. Therefore we translate the CLC classes to the land use classes used by
474 the Copernicus land cover map. Some examples of changes include: changing from Dump sites into
475 Urban fabric is classified as “*No change*”, changing from Non-irrigated arable land into Urban fabric to
476 “*Urbanization*”, changing from Airports to Mineral extraction sites to “*Other*” etc.

477 **Prevalent change**

478 We also mapped prevalent change on a 5×5 km grid and the change intensity on 20×20 km grid. We
479 divide the entire area of Europe into 5×5 km grids and count the number of pixels for each change class
480 within these blocks. The change class that covers the biggest amount of pixels is then assigned to the
481 corresponding pixel / grid node. We also save the number of pixels that the most prevalent change class
482 covers, as part of the total of pixels in the area for 20×20 km areas. For example: we used 30×30 m
483 resolution data, in each 20×20 km block we have $(20,000/30) \cdot (20,000/30) = 444,444$ pixels. If the
484 prevalent change class covers >94,000 pixels this means that it covers >20% of the total area.

Table 6. Harmonization scheme used to convert **ODSE-LULC** nomenclature to Copernicus Global Land Cover classes. On the left side, **ODSE-LULC** classes are converted to Forest, Other Vegetation, Wetland, Bare, Cropland, Urban, and Water classes. Each transition from one Copernicus class to another is then categorized into a change class in the cross-table.

ODSE-LULC class	Copernicus change class	Forest	Other Vegetation	Wetland	Bare	Cropland	Urban	Water
311: Broad-leaved forest 312: Coniferous forest	Forest			Deforestation		Deforestation and crop expansion	Deforestation and urbanization	
321: Natural grasslands 322: Moors and heathland 324: Transitional woodland-shrub 323: Sclerophyllous vegetation	Other Vegetation				Other	Desertification	Crop expansion	Urbanization
411: Inland wetlands 421: Maritime wetlands	Wetland			Wetland degradation		Wetland degradation and desertification	Wetland degradation and crop expansion	Wetland degradation and urbanization
332: Bare rocks 333: Sparsely vegetated areas 334: Burnt areas 335: Glaciers and perpetual snow 335: Beaches, dunes, and sands	Bare			Other			Crop expansion	
211: Non-irrigated arable land 212: Permanently irrigated arable land 213: Rice fields 221: Vineyards 222: Fruit trees and berry plantations 223: Olive groves 231: Pastures	Cropland	Reforestation		Land abandonment		Land abandonment and desertification		Urbanization
111: Urban fabric 122: Road and rail networks and associated land 123: Port areas 124: Airports 131: Mineral extraction sites 132: Dump sites 133: Construction sites 141: Green urban areas	Urban		Other					
511: Water courses 512: Water bodies 523: Sea and ocean 522: Estuaries 521: Coastal lagoons	Water		Water reduction					

RESULTS**Comparison of training data to other land cover products**

Table 7 provides an overview of each compared land cover product's accuracy when validated on subsets of ODSE-LULC training data. The S2GLC product scored the highest in both 2016 and 2018 subsets of our dataset, while the land cover product made by Pflugmacher et al. (2019) fits more closely to the 2015 subset. The 2019 point subset was considered too small to perform any meaningful comparison between ELC10 and GLC FCS30.

Table 7. Weighted F1-score of other land cover products when validated with the ODSE-LULC training dataset.

Land cover product	Validation year	Data source	Samples	Weighted F1-Score	Number of classes	Res. (m)
S2GLC	2016	LUCAS	756	0.724	8	10
Pflugmacher et al. (2019)	2016	LUCAS	719	0.719	10	30
GLC FCS30–2015	2016	LUCAS	724	0.677	10	30
Pflugmacher et al. (2019)	2015	LUCAS	144,027	0.657	11	30
S2GLC	2018	LUCAS	295,152	0.653	11	10
S2GLC	2018	CLC	1,000,063	0.604	12	10
ELC10	2018	LUCAS	42,629	0.596	8	10
GLC FCS30–2015	2015	LUCAS	138,342	0.503	12	30
ELC10	2018	CLC	172,382	0.456	8	10
GLC FCS30–2020	2018	LUCAS	308,838	0.424	12	30
GLC FCS30–2020	2018	CLC	1,026,914	0.420	12	30

Spatiotemporal model

The result of the EML model optimization resulted in the following hyperparameters:

- Random forest: Number of trees equal to 85, maximum depth per tree equal to 25, number of covariates to find the best split equal to 89, and 20 as minimum number of samples per leaf.
- Gradient boosted trees: Number of boosting rounds equal to 28, maximum depth per tree equal to 7, minimum loss reduction necessary to split a leaf node equal to 1, L1 regularization term on weights equal to 0.483, learning rate equal to 0.281, greedy histogram algorithm to construct the trees, and softmax as objective function.
- Artificial Neural Network: Four fully connected hidden layers with 64 artificial neurons each; ReLU as activation function, dropout rate equal to 0.15 and batch normalization in all the layers; softmax as activation function for output layer; batch size and number of epochs equal to 64 and 50, respectively; and Adam with Nesterov momentum as optimizer considering 5e-4 as learning rate.
- Logistic Regression: SAGA solver and multinomial function to minimize the loss.

The variable importance, generated by the tree based algorithms and presented in Fig. 5, shows that the 50th quantile of Landsat green band is the most important covariate, derived on summer and fall. In addition to spectral bands, several Landsat spectral indices (NDVI, SAVI, MSAVI, NBR, NB2 and

508 NDWI) appeared among the 50 most important covariates, while the global surface water frequency was
 509 considered the second most important for Random Forest and the 7th for GBT. This proves that variations
 510 of NDVI data have significant impact on the predictive power of the models ranking as the 8th, 9th and
 511 10th most important variables for RF models and 3rd most important variable for the GBT models (Fig. 5).

512 The geometric temperatures and terrain covariates were ranked more important for Random Forest,
 513 with the slope and elevation in 6th and 7th ranking position, respectively, while for the gradient boosted
 514 trees only the slope appears in the result of the analysis, in the 26th position of the ranking. These
 515 differences in the variable importance indicate that a wider usage of the feature space could lead to a
 516 better prediction power for the EML model, compared to using single learners.

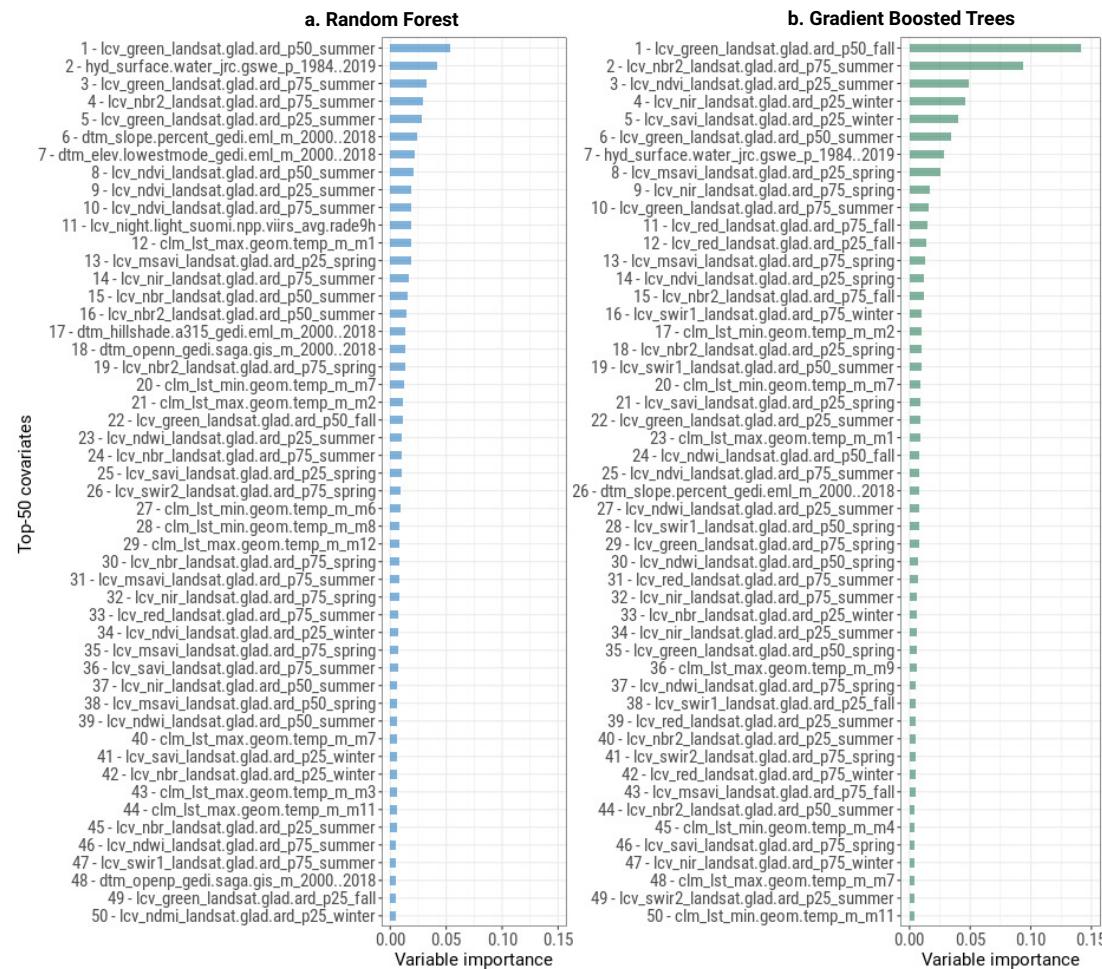


Figure 5. Variable importance for the top-50 covariates according to the (a) Random Forest and (b) Gradient Boosted Trees. The covariate names are composed by theme: land cover (lcv), digital terrain model (dtm), climate (clm) and hydrology (hyd); variable code: spectral bands (green, red, nir, swir1, and, swir2), spectral indices (ndvi, savi, msavi, nbr, nbr2, and ndwi), land surface temperature (lst), and openness positive negative derived (openp and openn). For the Landsat covariates the 25th, 50th and 75th quantiles are represented by p25, p50 and p75, respectively.

517 **Accuracy assessment of spatiotemporal models**

518 **Spatial 5-fold cross-validation**

519 The results of accuracy assessment using 5-fold cross-validation show that the ensemble achieved 62%
520 accuracy for 33 CLC level 3 classes, 70% accuracy for 14 level 2 classes, and 87% accuracy for the 5 level
521 1 classes (see also Table 9). Overall these results are consistent with similar studies by Calderón-Loor
522 et al. (2021) and Tsendlbazar et al. (2018).

523 Most false positive errors were recorded for classes 211: Non-irrigated arable land (16%), 324:
524 Transitional Woodland-shrub (14.8%), 311: Broad-leaved forest (14.2%) and 231: Pastures (13.7%).
525 Most false negative errors were for the same classes in a different order, and : 324 (17.9%), 311 (9.5%),
526 231 (8.7%), and 211 (8.6%). Class 1: Artificial surfaces had the lowest precision (0.78) and recall (0.65)
527 scores of all level 1 classes. The relatively low recall value is mainly due to false positives for class 211:
528 Non-irrigated arable land. Its most accurately predicted subclass was 111: Urban fabric (0.68). This class
529 had a precision of 59% and a positive rate of 1.38; the main cause are false positives on other classes
530 inside its level 1 class. Class 2: Agricultural areas had a precision of 0.81 and a recall of 0.9. The main
531 causes of this relatively low precision score are false positives for class 211: Non-irrigated arable land
532 in non-forested, non-water body classes, and for class 231: Pastures in sub-classes of class 3: Forests
533 and seminatural areas. Class 3 had the highest number of classes and the largest support in the training
534 data. The largest cause for error within the class are confusions between 311: Broad-leaved forest, 312:
535 Coniferous forest, and 324: Transitional woodland-shrub. The highest accuracy (0.93) was achieved for
536 class 512: Water bodies. This class and 421: Maritime Wetlands were the main false positive class for
537 other open water classes (511, 521, 522, 523). Table 8 shows the spatial cross-validation performance
538 separated per year when assessed on full thematic resolution. The average accuracy per year was 0.61,
539 with a standard deviation of 0.035.

Table 8. Spatial cross-validation performance (Weighted F1-score) of our ensemble model per year in the training data.

Year	Number of points	Weighted F1-score
2000	1.034.449	0.652
2006	1.228.471	0.645
2009	225.515	0.585
2012	1.333.239	0.637
2015	147.554	0.589
2016	807	0.562
2018	1.392.044	0.628
2019	149	0.562
Average		0.607
Standard deviation		0.035

540 **Independent test data**

541 Overlaying our 2017 LULC predictions with the S2GLC dataset, reclassing them to the S2GLC nomen-
542 clature, and removing any points with class 412: Peatbogs resulted in 51,279 data points. 2914 points had
543 a predicted class that was not in the S2GLC nomenclature (see Table 5). The ‘conservative’ assessment

Table 9. Classification report for three levels of thematic resolution: 33 classes (level 3), 14 classes (level 2), and 5 classes (level 1)

Class Level 1	Level 2	Level 3	Precision			Recall			F1-score			Support		
			L3	L2	L1	L3	L2	L1	L3	L2	L1	L3	L2	L1
1: Artificial surfaces	11: Urban fabric	111: Urban fabric	0.59	0.59		0.82	0.82		0.68	0.68		146459	146459	
	12: Industrial, commercial and transport units	122: Road and rail networks and associated land	0.37			0.06			0.10			46173		
		123: Port areas	0.49	0.45		0.08	0.07		0.13	0.12		16733	69507	
		124: Airports	0.29			0.04			0.07			6601		
	13: Mine, dump and construction sites	131: Mineral extraction sites	0.51		0.78	0.35		0.65	0.41		0.71	37116		308799
		132: Dump sites	0.35	0.57		0.02	0.26		0.03	0.35		6745	59637	
2: Agricultural areas	14: Artif. non-agri. veg. areas	133: Construction sites	0.20			0.02			0.03			15776		
		141: Green urban areas	0.25	0.25		0.12	0.12		0.16	0.16		33196	33196	
	21: Arable land	211: Non-irrigated arable land	0.66			0.79			0.72			830748		
		212: Permanently irrigated arable land	0.42	0.69		0.19			0.26	0.74		32637		
		213: Rice fields	0.72			0.15	0.80		0.24			14840	878225	
	22: Permanent crops	221: Vineyards	0.52		0.81	0.36		0.90	0.43		0.85	51027		1652724
3: Forests and seminatural areas	222: Fruit trees and berry plantations	0.38				0.14			0.21			49432		
	223: Olive groves	0.45	0.59			0.44	0.41		0.44	0.48		49628	150087	
	23: Pastures	231: Pastures	0.62	0.62		0.72	0.72		0.66	0.66		624412	624412	
	31: Forest	311: Broad-leaved forest	0.67	0.78		0.75	0.83		0.71	0.81		773914	1440408	
		312: Coniferous forest	0.74			0.74			0.74			666494		
	32: Scrub and/or herbaceous vegetation associations	321: Natural grasslands	0.38			0.30			0.34			209897		
4: Wetlands	322: Moors and heathland	0.49	0.65			0.33	0.57		0.40	0.61		263842		
	323: Sclerophyllous vegetation	0.32				0.39			0.35	0.61		132151	1265286	
	324: Transitional woodland-shrub	0.49		0.93		0.45		0.88	0.47		0.90	659396		2973800
	33: Open spaces with little or no vegetation	331: Beaches, dunes, sands	0.51			0.23			0.31			22392		
		332: Bare rocks	0.65			0.39			0.49			64101		
		333: Sparsely vegetated areas	0.51	0.72		0.43	0.52		0.47	0.60		162694	268106	
5: Water bodies	334: Burn areas	0.24				0.00			0.00			12036		
	335: Glaciers and perpetual snow	0.85				0.82			0.84			6883		
	41: Inland wetlands	411: Inland wetlands	0.70	0.70	0.70	0.73	0.73	0.73	0.71	0.71	0.72	208087	208087	223490
	42: Maritime wetlands	421: Maritime wetlands	0.62	0.62		0.66	0.66		0.64	0.64		15403	15403	
	51: Inland waters	511: Water courses	0.29			0.08			0.13			12830		
		512: Water bodies	0.91	0.92		0.96	0.93		0.93	0.93		185788	198618	
	521: Coastal lagoons	0.51		0.93		0.35		0.92	0.41		0.93	2597		203416
	52: Maritime waters	522: Estuaries	0.37	0.60		0.11			0.17	0.41		1720		
		523: Sea and ocean	0.52			0.22	0.31		0.31			481	4798	
	Accuracy		0.62	0.70	0.87	0.62	0.70	0.87	0.62	0.70	0.87			
	Macro average		0.50	0.63	0.83	0.37	0.55	0.82	0.39	0.57	0.82			
	Weighted average		0.60	0.70	0.87	0.62	0.70	0.87	0.60	0.69	0.87		5362229	

544 (on 50,939 points) including the non-S2GLC classes resulted in a weighted F1-score of 0.877, a macro
545 F1-score of 0.741 and a kappa score of 0.828. The '*optimistic*' assessment excluding non-S2GLC predic-
546 tions resulted in a weighted F1-score of 0.903, a macro F1-score of 0.772 and a kappa score of 0.882 (see
547 Table 10).

548 Taking into account possible noise from the translation process, these results are similar to those
549 reported by Malinowski et al. (2020). Weighted precision, recall and F1-scores are also higher than our
550 cross-validation scores at all thematic resolution levels (see Table 9).

Table 10. Optimistic classification report of our 2017 LULC prediction on 48,365 S2GLC points after removing 2914 points with predicted classes without an equivalent S2GLC class (141: Green urban areas, 222: Fruit trees and berry plantations, 223: Olive groves, 324: Transitional woodland-shrub, 333: Sparsely vegetated areas, and 334: Burnt areas).

S2GLC Class	Precision	Recall	F1-Score	Support
Artificial surfaces	0.923	0.891	0.906	1821
Cultivated areas	0.914	0.946	0.930	13404
Vineyards	0.771	0.754	0.762	479
Herbaceous vegetation	0.798	0.811	0.805	6486
Broadleaf tree cover	0.959	0.964	0.962	9956
Coniferous tree cover	0.965	0.967	0.966	8523
Moors and heathland	0.747	0.618	0.676	1484
Sclerophyllous vegetation	0.686	0.346	0.460	619
Natural material surfaces	0.884	0.864	0.874	1821
Permanent snow	0.544	0.840	0.660	81
Marshes	0.234	0.548	0.328	305
Water bodies	0.996	0.876	0.932	3386
				48365
Macro average	0.785	0.785	0.772	
Weighted average	0.907	0.903	0.903	

551 **Comparison of spatial and spatiotemporal models**

552 Table 11 shows the weighted F1-scores of spatial models trained on 100,000 points sampled from one year,
553 spatiotemporal models trained on 100,000 points sampled from across multiple years, and spatiotemporal
554 models trained on 100,000 points per year. Each model was validated on 33,333 points from the same
555 year(s) as its training data, and on 33,000 points from the year 2018, which was left out of all training
556 datasets.

557 Results show that all models performed better when classifying in known years, regardless of data
558 source. The spatiotemporal model trained on only CLC points achieved the highest F1-scores for both
559 known-year and unknown-year classification. This model outperformed spatial models on known-year
560 classification by 2.7% and unknown-year classification by 3.5% as seen in Table 11.

561 Fig. 6 shows that spatial models achieved a higher F1 score than spatiotemporal models when
562 classifying LULC on the validation set from the year they were trained on, and that spatiotemporal models
563 outperformed spatial models when generalizing to the 2018 validation datasets.

Table 11. Trained year and untrained year weighted F1-scores of spatial and spatiotemporal models trained on CLC points, LUCAS points, and a combination of both.

Model	CLC		LUCAS		CLC+LUCAS	
	Trained year(s)	2018 (untrained)	Trained year(s)	2018 (untrained)	Trained year(s)	2018 (untrained)
Spatial - 2000			0.610	0.542	0.611	0.515
Spatial - 2006	0.595	0.437	0.604	0.563	0.587	0.534
Spatial - 2009	0.595	0.482			0.602	0.415
Spatial - 2012	0.559	0.476	0.611	0.574	0.565	0.529
Spatial - Average	0.583	0.465	0.608	0.560	0.591	0.498
Spatiotemporal - 100.000 pts.	0.612	0.576	0.568	0.478	0.574	0.532
Spatiotemporal - 100.000 pts. per year	0.625	0.579	0.608	0.491	0.595	0.543

Weighted F1 score of spatial and spatiotemporal models with varying training dataset size

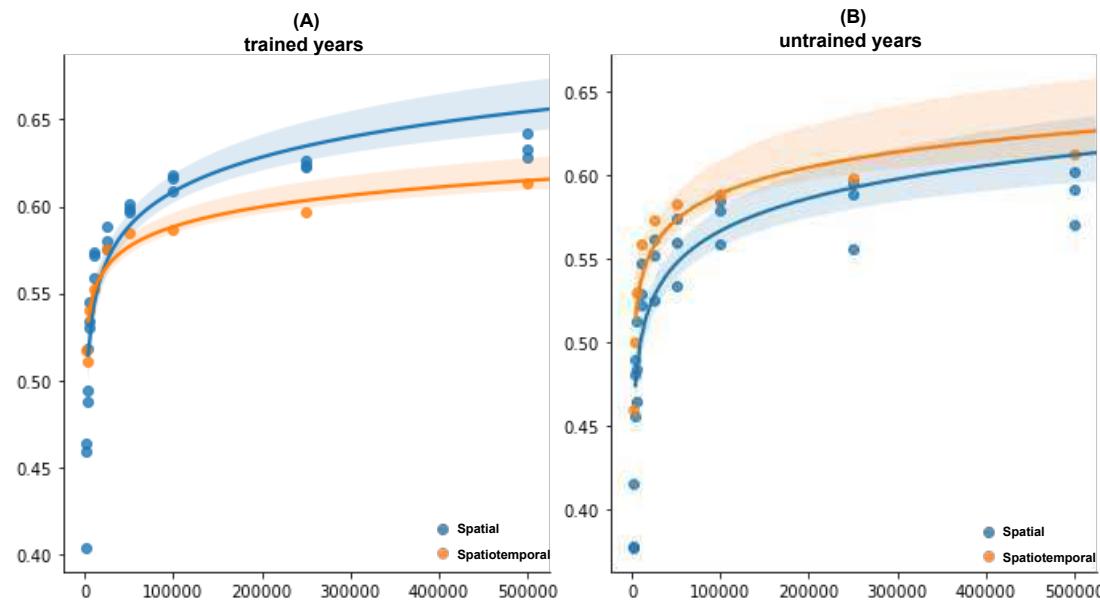


Figure 6. Performance comparison of spatial and spatiotemporal models when classifying LULC with varying training dataset sizes on (A): data from years that were included in their training dataset, and (B): data from 2018, which was excluded from all training data for this experiment

564 Results Time-series analysis

565 NDVI and land use class probability slopes

566 Our NDVI slope maps show which areas have an increase or decrease in NDVI over time. When reviewing
567 the maps big features as well as details can be appreciated as seen in Fig. 4.

568 Fig. 4.1 and 4.2 show areas of negative and positive slope occur adjacent to each other without gradual
569 transitions. Fig. 4.3 and 4.4 show examples of relatively large areas with homogeneous NDVI slope
570 values. Overall, NDVI slopes in Europe tend to be positive, the largest exceptions being negative slope
571 regions in Northern Scandinavia, Scotland, the Alps, South West France, Spain, Italy and Greece. The
572 NDVI slope patterns in Sweden mostly match probability slopes for coniferous forests.

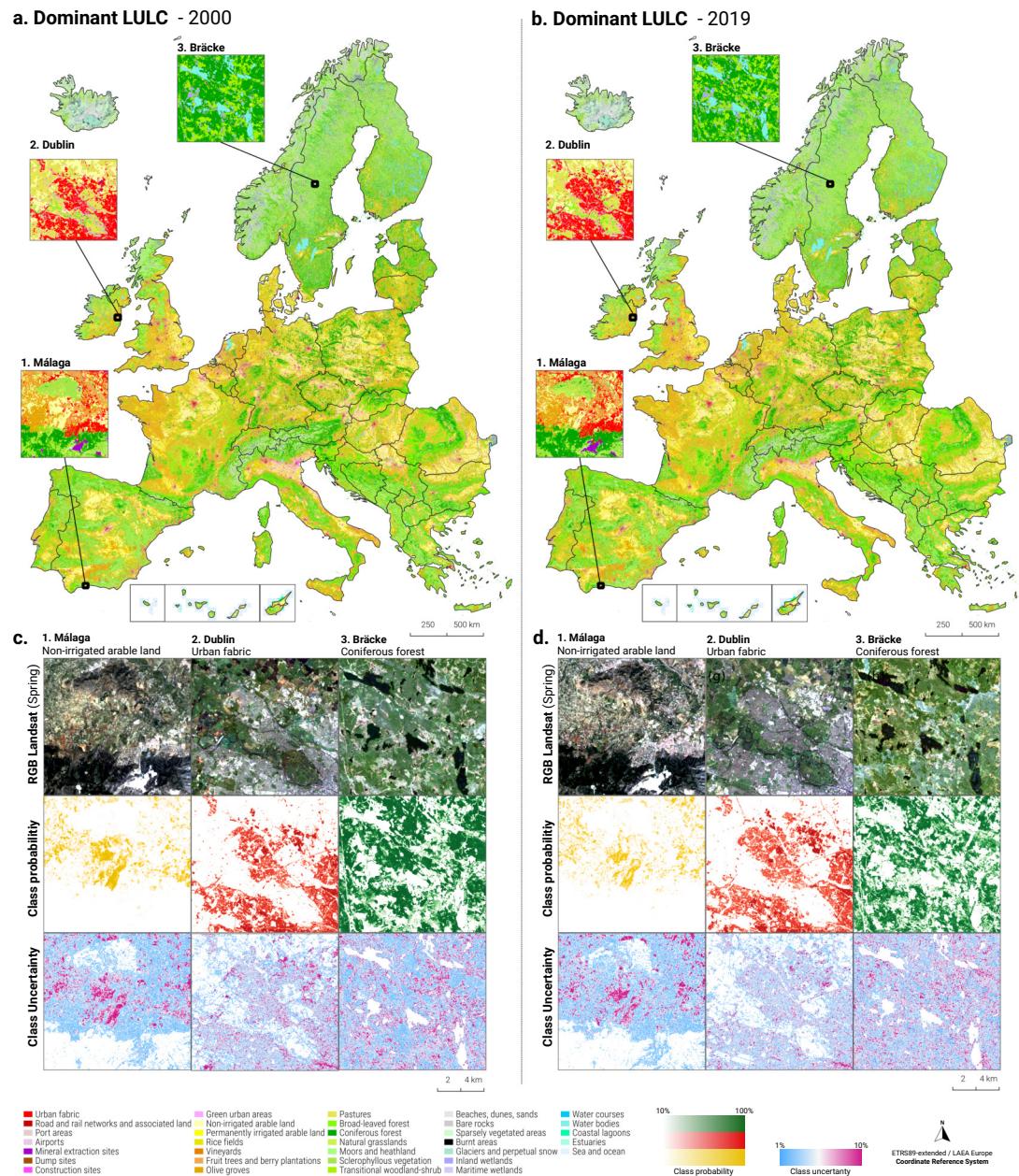


Figure 7. Dominant LULC classes, predicted probability and uncertainty for Non-irrigated arable land, Coniferous forest and Urban Fabric, RGB Landsat temporal composite (Spring season) for the years 2000 and 2019.

Table 12. Confusion matrix of 5-fold spatial cross-validation predictions for 33 classes

	111	122	123	124	131	132	133	141	211	212	213	221	222	223	231	311	312	321	322	323	324	331	332	333	334	335	411	421	511	512	521	522	523	
111: Urban fabric	119591	1878	423	93	1797	77	247	1831	11203	146	0	634	587	1112	3199	971	567	457	94	280	902	176	1	78	1	1	31	30	21	29	0	2	0	
122: Road and rail ...	19743	2745	139	34	596	17	68	1242	9076	113	20	354	334	628	3727	2366	1699	507	84	552	1768	79	2	110	0	1	70	27	44	26	0	2	0	
123: Port areas	9266	374	1280	36	369	6	40	412	1981	37	0	103	120	201	853	332	145	132	13	71	210	144	2	25	0	0	6	215	66	222	38	33	1	
124: Airports	1338	80	33	272	123	1	28	35	1572	16	0	10	10	11	2249	78	46	208	50	31	260	64	0	54	0	0	21	6	4	1	0	0	0	
131: Mineral extraction	4553	149	54	11	12913	64	133	80	10664	189	17	274	133	291	832	319	222	544	107	611	1190	400	81	1173	2	0	196	50	196	1668	0	0	0	
132: Dump sites	1717	139	29	8	1129	123	37	99	1424	9	0	38	10	20	647	285	84	114	26	60	291	30	3	57	0	0	53	16	43	254	0	0	0	
133: Construction sites	2360	83	48	17	698	1	296	65	8846	220	17	177	118	198	728	229	155	383	43	280	477	82	3	138	2	0	38	34	12	28	0	0	0	
141: Green urban areas	7048	296	5	25	27	5	7	3946	4070	28	0	146	280	229	6829	6376	2225	154	115	59	1092	4	0	7	0	0	81	6	50	86	0	0	0	
211: Non-irrigated arable ...	9446	285	19	97	1602	14	149	995	656020	4138	280	4592	1461	2841	120757	5105	975	7134	377	2701	9345	276	0	312	4	0	905	139	63	716	0	0	0	
212: Permanently irrigated ..	1413	31	3	4	112	0	44	2	19063	6271	156	1006	882	690	982	54	424	27	226	402	91	0	60	0	0	8	25	14	39	0	0	0		
213: Rice fields	129	9	1	0	29	0	5	15	10141	445	2157	122	70	51	860	264	43	72	9	30	227	30	0	2	0	0	53	58	6	11	1	0	0	
221: Vineyards	1881	36	1	2	147	0	9	85	15648	425	9	18455	1359	2667	3255	2321	297	394	33	853	3032	25	1	51	1	0	8	4	8	20	0	0	0	
222: Fruit trees and berry ...	2251	55	1	1	133	0	35	534	11977	638	18	2131	7117	4051	6402	7252	397	721	34	1693	3711	38	0	159	0	0	28	10	15	30	0	0	0	
223: Olive groves	1549	13	0	0	51	0	16	11	6802	107	2	1558	1776	21813	869	5144	620	1106	37	5110	2957	24	0	50	0	0	0	2	7	4	0	0	0	
231: Pastures	5284	141	16	159	711	8	87	1363	100674	198	13	1120	413	909	447551	14011	2120	18502	2675	2317	20654	149	4	483	1	0	3245	316	116	1166	3	1	2	
311: Broad-leaved forest	1453	127	2	3	110	3	1	2179	10103	158	12	731	869	2046	18825	580221	55928	2000	11770	11490	70372	28	102	562	3	0	4071	48	142	554	1	0	0	
312: Coniferous forest	904	97	3	0	66	2	2	604	3308	17	0	156	202	1207	4367	68627	495423	1212	6375	9008	67833	16	8	1025	3	0	5506	23	20	480	0	0	0	
321: Natural grasslands	3449	206	15	89	378	3	30	758	20695	155	4	451	364	1228	41217	4955	1802	63942	13172	18280	22418	243	557	11600	5	0	3274	286	36	285	0	0	0	
322: Moors and heathland	2017	133	7	32	295	1	21	477	16432	159	7	610	580	1619	15162	26323	10442	17948	88169	14335	30574	206	1244	20878	17	21	14543	298	35	1249	1	2	5	
323: Sclerophyllous ...	737	51	4	2	302	0	34	2	5997	89	1	266	396	2591	1839	16792	6983	14013	1394	51736	26085	119	14	2472	11	0	8	108	28	69	8	0	0	0
324: Transitional wood...	2954	193	10	19	771	11	39	787	31865	368	26	1463	836	2704	32856	115581	78257	17257	17146	30915	294848	191	149	5793	21	0	23093	173	113	946	1	0	10	
331: Beaches, dunes, sands	1219	60	138	15	909	1	23	48	7145	298	74	318	127	323	847	438	204	676	268	401	756	5052	479	1088	0	0	206	759	215	290	14	1	0	
332: Bare rocks	533	27	14	1	349	1	27	35	6521	278	73	310	85	307	422	617	738	2196	2093	1738	2584	521	25123	18147	0	743	242	116	22	229	1	1	7	
333: Sparsely vegetated ...	927	31	18	16	754	4	46	52	7919	302	73	315	184	552	1694	2769	2412	14411	23807	7531	12392	666	9504	70514	1	209	4805	192	34	547	4	1	8	
334: Burnt areas	347	11	4	0	162	0	26	27	7421	162	17	158	83	186	405	337	244	378	219	563	1047	21	1	136	23	0	39	11	2	6	0	0	0	
335: Glaciers and ... snow	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	7	0	0	28	1049	85	0	5663	1	0	11	38	0	0	0	
411: Inland wetlands	76	24	0	0	50	0	0	47	1689	32	12	36	43	10	6765	3951	3878	2066	12760	13	19623	81	21	2030	0	0	151249	716	149	2764	2	0	0	
421: Maritime wetlands	62	19	102	1	29	0	0	14	295	6	6	5	27	1	1106	161	63	237	231	34	132	510	1	74	0	0	1173	10137	120	506	236	108	7	
511: Water courses	376	112	58	5	361	3	0	111	809	27	17	23	89	52	653	1271	443	47	47	82	322	447	16	74	0	2	733	492	1070	4955	80	53	0	
512: Water bodies	168	52	55	0	421	4	2	57	527	10	3	10	9	559	703	732	21	170	26	290	76	83	147	0	16	2080	476	898	177837	263	47	36		
521: Coastal lagoons	3	1	61	0	1	0	0	0	7	0	0	0	0	0	13	2	7	3	7	2	6	42	1	3	0	0	54	711	31	665	898	64	15	
522: Estuaries	2	1	77	0	2	0	0	1	6	0	0	0	0	0	6	6	1	4	9	1	0	32	1	2	0	0	25	748	77	357	163	189	10	
523: Sea and ocean	2	0	12	0	1	0	0	0	3	0	0	0	0	0	3	2	2	7	4	1	1	9	0	1	0	0	7	119	9	152	31	7	108	

573 **Land use change classes**

574 We generated yearly maps for hard change classes. Filtered data as well as the removed noise can be
 575 viewed from the ODS-Europe viewer. The right-most subplots of Fig. 4 show examples of where sudden
 576 land cover change classes at 30×30 m tend to match relatively large negative slopes, especially for change
 577 classes such as deforestation and urbanization.

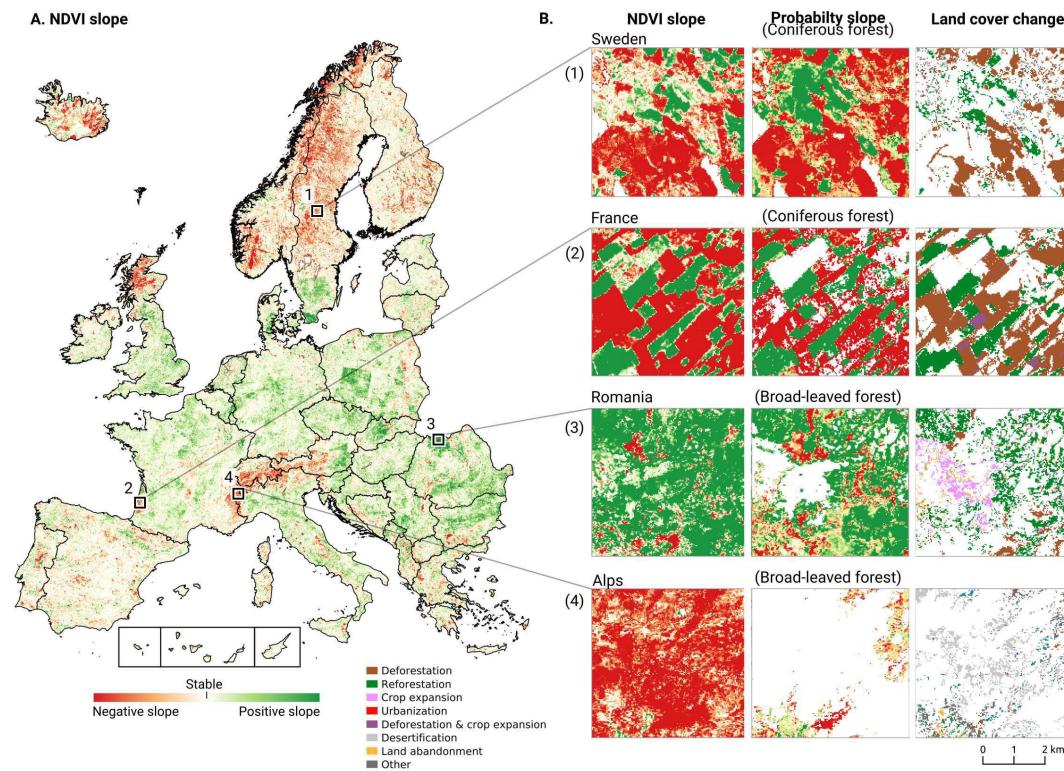


Figure 8. Trends in NDVI values between 2000 and 2019 compared to trends in probabilities from machine learning and hard change classes between 2001 and 2018.

578 We visualized the dominant type of change in a 5×5 km grid in Fig. 9. We also present the intensity of
 579 change as part of the total area on a separate map using 20×20 km areas (Fig. 9). Large parts of mainland
 580 Europe are characterized with reforestation as the main change with patches of urbanization scattered in
 581 between. Norway, Sweden and Finland are characterized with deforestation as the main land use change
 582 class. Large areas in Spain have land abandonment and crop expansion as the main land use class. When
 583 taking into account the intensity of the changes the central European countries seem to be fairly stable
 584 with the Iberian peninsula, Scandinavia and parts of eastern Europe exhibiting more intense changes.

585 **DISCUSSION**

586 **Summary findings**

587 We presented a framework for automated prediction of land cover / land use classes and change analysis
 588 based on spatiotemporal Ensemble Machine Learning and per-pixel trend analysis. In this framework
 589 we focused not only on predicting the most probable class, but also on mapping each probability and
 590 associated uncertainty. Such detailed information allows any future users to limit the decisions based

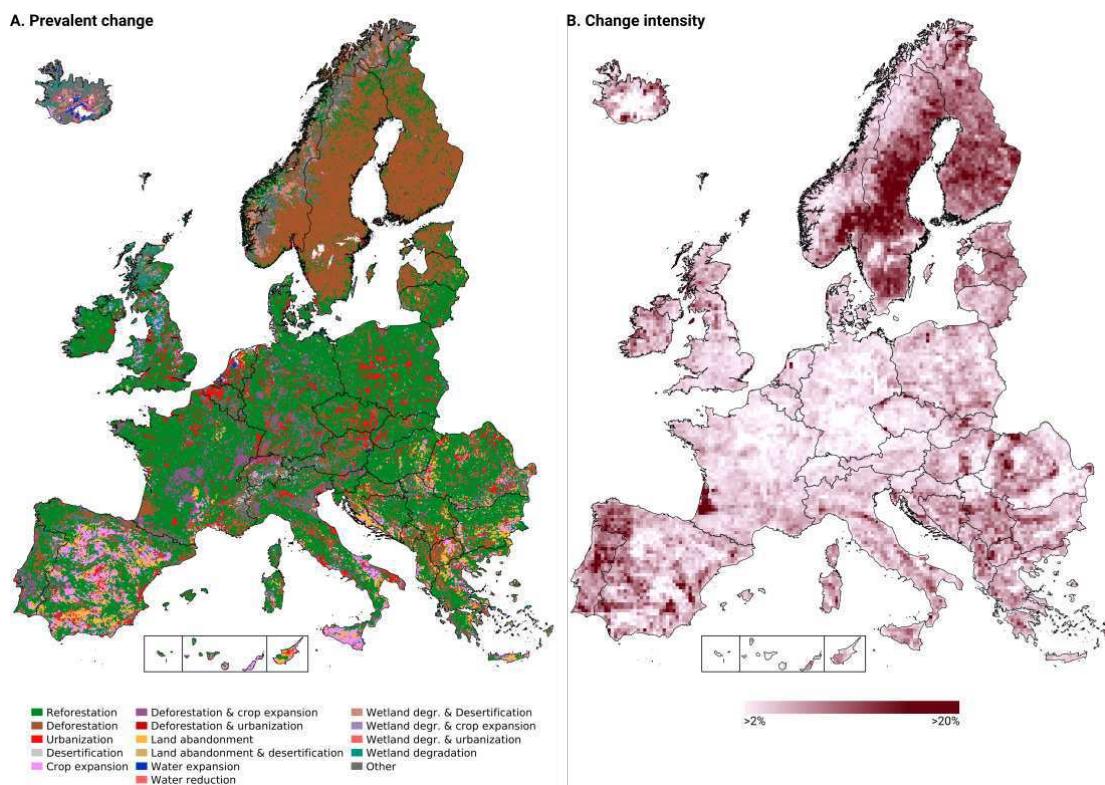


Figure 9. Prevalent change map at 5×5km tiles and change intensity at 20×20km tiles.

591 on uncertainty per pixel and/or incorporate it in further spatial modeling. Producing such detailed data,
592 however, comes at a significant cost with the output dataset about 100-times larger (currently about
593 10 TiB of data representing land cover and land use for Europe at relatively fine spatial resolution of 30 m)
594 than produced by similar European land cover mapping initiatives (Pflugmacher et al., 2019; Malinowski
595 et al., 2020; Venter and Sydenham, 2021).

596 We further explained the time-series analysis framework for processing partial probabilities and
597 NDVI values aiming at detection of significant spatiotemporal trends. We provide pixel-wise uncertainty
598 measures (standard deviation of the slope / beta coefficient and R-square), which can also be used in
599 any further spatial modeling. The whole framework, from hyper-parameter optimisation, fine-tuning,
600 prediction and time-series analysis, is fully automated and generates consistent results over time with
601 quantified uncertainty, making it more cost-effective for future updates and additions.

602 To enable easy access to the data and collaboration on similar research projects, we have made all
603 the data, including training points and classification matrices reported in this paper, and code used to
604 fine-tune the models and produce predictions, available via the project repository https://gitlab.com/geoharmonizer_inea; and all maps available via the Open Data Science Europe (ODS-Europe)
605 viewer at <https://maps.opendatascience.eu> (Fig. 10). Most of the layers available via the ODS-
606 Europe viewer are also available seamlessly through our S3 Cloud Object Service as Cloud-Optimized
607 GeoTIFFs under the Open Data Commons Open Database License (ODbL) and/or Creative Commons
608 Attribution-ShareAlike 4.0 and/or Creative Commons Attribution 4.0 International license (CC BY). This
609 means that you can (a) visualize the data and run processing directly using QGIS or similar, (b) import,

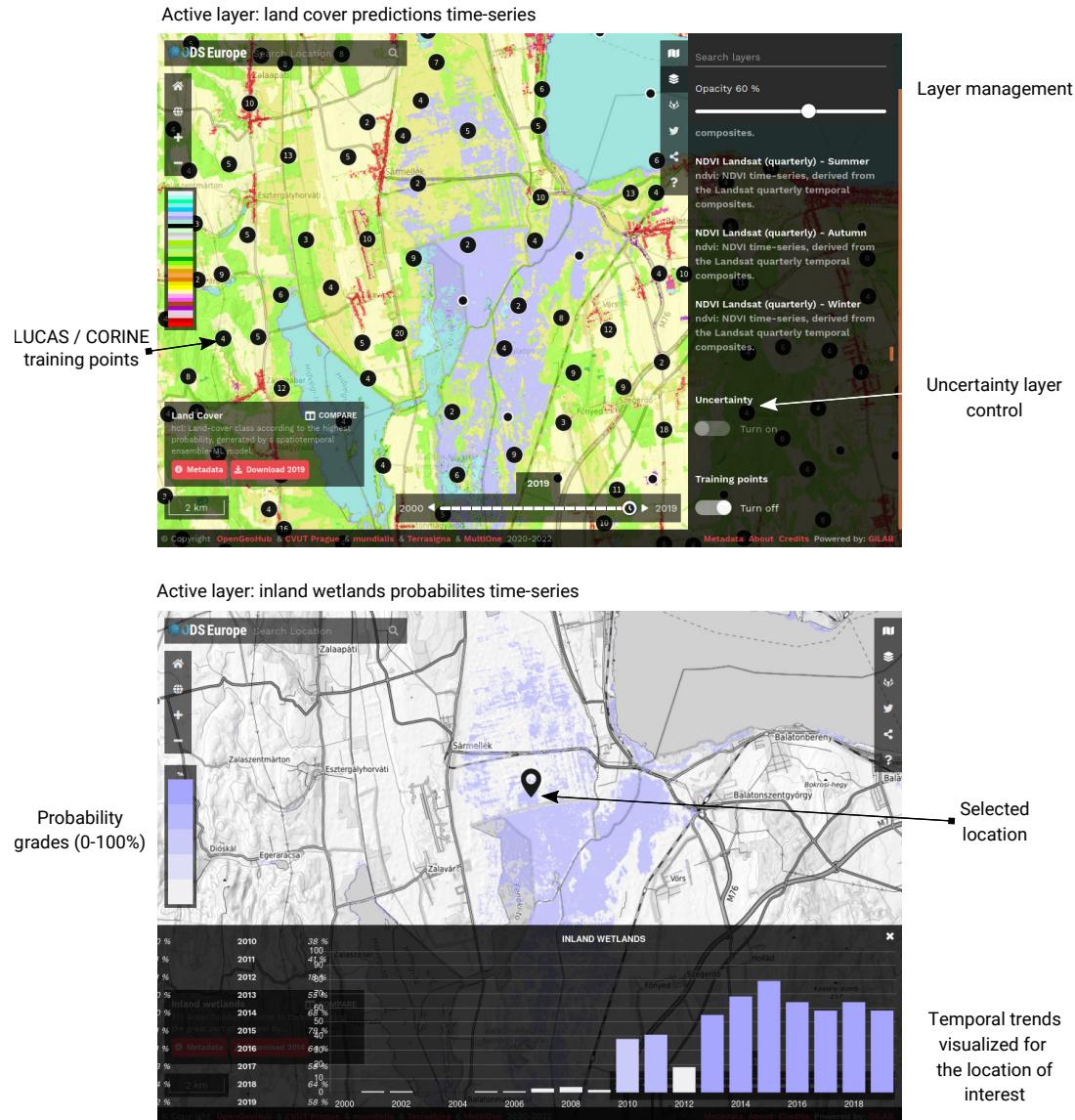


Figure 10. Visualization of the ODSE-LULC predictions and training points used to produce time-series of land cover maps: the Open Data Science Europe viewer and its main functionality. The viewer supports visualization of data in 2D (OpenLayers) and 3D (Cesium).

611 subset, crop and overlay parts of the data for a local area. We do not however recommend downloading
612 whole data sets by using Wasabi.com. To download the complete datasets (the whole of the European
613 continent) we recommend using zenodo.org or similar.

614 Our accuracy assessment results indicate limited mapping accuracy (62%) at highest classification
615 level (33 classes) with several classes such as “*airports*”, “*burned areas*” performing poorly, effectively
616 being on the edge of usable. Noise in predictions is visible also by scrolling through time-lapse animations
617 via the data portal (<https://maps.opendataservice.eu>), with some pixels being classified into
618 transitional class sequences that are highly unlikely or physically impossible — for example a pixel
619 changing into coniferous forest class, then to urban area, then next year again into forest class.

620 Our comparison of spatial versus spatiotemporal ML indicates that there is an added value in using
621 spatiotemporal ML, especially for predicting land cover for years before, beyond or between different
622 training point campaigns: the spatiotemporal model outperforms spatial models on known-year classi-
623 fication by 2.7% and unknown-year classification by 3.5% (Fig. 6). Also, we have demonstrated that
624 spatiotemporal ML allows us to remove any bias in predictions through time (Table 8), which justifies
625 further implementing time-series analysis on the produced time-series of predictions of hard classes,
626 probabilities and remote sensing indices (NDVI).

627 **Comparison with other land cover products for Europe**

628 Validating our 2017 classification with the **S2GLC** test dataset resulted in similar performance values
629 as reported by **Malinowski et al. (2020)**, and higher than our cross-validation performance: results of
630 the accuracy assessment using 48,365 independent test samples shows 87% match with the validation
631 points. This indicates the nomenclature used by **Malinowski et al. (2020)** is possibly more suitable for
632 remote sensing-based classification. It also shows that accuracy of the **ODSE-LULC** is comparable even
633 with 10 m resolution products, and most importantly — with some improvements in accuracy — it could
634 potentially in near future match the 85% accuracy threshold required by the **CLC** project.

635 **Advantages and limitations of using spatiotemporal EML**

636 The spatiotemporal ML described in this paper comes with multiple advantages:

- 637 • A single spacetime model can be used to model phenomena such as land cover — making it a
638 holistic approach to data harmonization and prediction that performs consistently across multiple
639 years;
- 640 • By including **LUCAS** points we base modeling and predictions on a consistent and quality-
641 controlled data set that allows for unbiased assessment of land cover dynamics;
- 642 • In principle, we can also use the existing model to predict land cover for years 2020 and 2021
643 without collecting new training data, as preparing Landsat images for these periods would be likely
644 enough;
- 645 • As all processes shown above are fully automated we can continue improving the models and
646 re-running predictions with relative ease.

647 Possibly the biggest advantage of using spacetemporal ML is that it allows for using [LUCAS](#) and [CLC](#)
648 data to make consistent predictions for periods prior to year 2000 for which very little training data is
649 available. In the next phase of the project we will focus on producing predictions for years 1995, 1990
650 and to 1985. Currently we are not in position to make predictions for these years due to the Landsat
651 [ARD](#) product ([Potapov et al., 2020](#)) being available only for periods >1997. We would need to compute
652 ourselves and recalibrate the Landsat products, which added a higher level complexity also because of
653 the differences between Landsat 5, 6, 7 and 8 sensors.

654 Some obvious limitations of our approach include:

- 655 • High computational intensity / still relatively high costs of preparing, producing and storing large
656 amounts of data.
- 657 • Limited spatial resolution of 30 m, where currently most of land cover products aim at 10 m
658 resolution.
- 659 • High dependence on quality training data spread through time. Many continents do not have even
660 close to similar datasets as [LUCAS](#) and hence transferring this methodology to other continents
661 (apart from USA and Australia) would likely be a cumbersome.
- 662 • Inconsistent accuracies per class result in a product with a variable quality. This can put many
663 potential users off as consistency is often crucial for usability of geo-data.

664 The spatial cross-validation results show that the ensemble did not perform consistently across all
665 classes at the highest level of thematic resolution. For instance, classes 111 and 323 had positive rates
666 of 1.38 and 1.22. Fig. 11 shows that many classes in the same level 1 category (1: Artificial surfaces)
667 were frequently incorrectly classified as class 111, while 211: Non-irrigated arable land — the level 3
668 class with the highest support was the leading false positive class for the largest number of classes. Poor
669 accuracy of some classes and high dependence of spatiotemporal ML on quality training data, could be
670 considered overall highest limitations of this work.

671 **Time-series analysis, interpretations and challenges**

672 The results of [NDVI](#) and probability trend analysis show some interesting patterns. We have focused on
673 four geographic areas: (1) Sweden, as its forest dynamics have already garnered academic attention and
674 it is an exemplary area where remote sensing techniques and on the ground measurements might come
675 to different conclusions (see e.g. [Ceccherini et al. \(2020\)](#)). (2) South West France, as it is similar to the
676 Sweden both in our data and compared and is also compared by other authors ([Senf and Seidl, 2021](#)). (3)
677 Northern Romania because it shows a large region with positive trends for both [NDVI](#) and broad-leaved
678 forest land cover, suggesting it is reforesting at high rates. Finally, we found large regions in the Alps (4)
679 show a strong negative trend for [NDVI](#) values that does not seem to correspond to an unbalanced harvesting
680 of forests or other obvious processes. This signal in our data is not yet understood and will need additional
681 research.

682 Intense forest loss is a well studied process in Europe ([Senf et al., 2018](#); [Ceccherini et al., 2020](#); [Senf
683 and Seidl, 2021](#)). Discrepancies between national forest inventories and remote sensing technique has led

684 to disagreements in Sweden (Paulsson et al., 2020), Finland (Korhonen, 2020), and Norway (Rossi et al.,
685 2019). For instance, it was found that existing remote sensing products are deemed not fit for these types
686 of analysis (Palahí et al., 2021). For these reasons, and because we do not validate our trend results, we
687 neither attribute specific causes, nor do we analyze differences between specific time periods.

688 We do compare the most prominent change between 2001–2018 and our results suggest that forest is
689 disappearing more than it is re-appearing in multiple locations. This is corroborated by Global Forest
690 Watch; for example, the Jämtland region in Sweden lost 287k ha of tree cover and gained 164k ha (Hansen
691 et al., 2013). We present the case of the Landes region in France here as well as it shows a similar pattern
692 to large parts of Sweden and is a known area for large scale forest harvesting (Senf and Seidl, 2021).
693 These cases exemplify the usefulness of our maps for finding similar processes all over Europe by using a
694 combination of the maps that are presented here.

695 Our data suggests that NDVI increase, accompanied by reforestation, is the mayor component of
696 change on a European scale. This is corroborated by the FAO's State of Europe's Forests report 2020
697 which states that European forest cover has increased by 9% between 1990 and 2020 (Raši, 2020) and
698 with global estimates that forest cover has increased by 7% between 1982 and 2016 (Song et al., 2018).
699 This increase is consistent with expectations that increased CO₂ will enhance plant growth in general but
700 increase in NDVI could also be explained by other factors such as increase in cropland cover (Huang et al.,
701 2018), having a higher NDVI value than other land use types, nitrogen depositions. Another concern
702 that is raised is that most of the increase in forest gain is by planted forests (Payn et al., 2015) that are
703 less valuable in terms of biodiversity and carbon sequestration (Liu et al., 2018) and less adaptable to
704 climate change. One exemplary area for reforestation is found in Northern Romania for all three parts of
705 our time-series analysis: Both NDVI and broad-leaved forest probability slopes show positive trends in a
706 large area marked with reforestation as the dominant change class.

707 Finally, our data for the Alps shows unexpected negative NDVI trends for large parts of the Alps. This
708 may be related to changes in snow cover as found by Wang et al. (2018) in the Tibetan Plateau and by
709 Buus-Hinkler et al. (2006) in the Arctic regions. However, this is not corroborated by the probability
710 slope for class "Glaciers and perpetual snow" in our data. It is also possible that this is an artifact from
711 our gap-filling step; further study is necessary before any conclusions can be drawn.

712 Future work

713 Even though our framework is comprehensive and has produced predictions of comparable accuracy to
714 the current state-of-the-art, after almost 14 months of processing the data and modeling land cover, we
715 have found that many aspects of our system could be improved:

- 716 • *Combining Landsat and Sentinel data:* We have not yet tested, even though we are aware of the
717 Harmonized Sentinel-Landsat product, using a combination of Sentinel 1/2 and Landsat imagery
718 for land cover mapping. This has proven to be rather complex, especially considering that our
719 interest is in mapping land cover for large time-spans — if possible going back to year 1984 —
720 while Sentinel data is only available for years >2016.
- 721 • *Cross-validation of land cover trends:* We have not yet put enough effort to independently identify

722 and quantify sudden changes in land cover due to natural hazards (fires, floods etc) but also due to
 723 slow inherent processes such as urbanisation, vegetation succession and similar. How well does our
 724 trend analysis matches the processes on the ground? This was beyond the scope of our project and
 725 we find it most important that the data is now available to any research groups interested in testing
 726 their usability for land monitoring projects.

727 • *Combining classification with Object-Based Image Analysis (OBIA) and pattern recognition:*
 728 Incorporating spatial context to our workflow could potentially improve performance for several
 729 classes that are defined by land use. For instance, class 124: “*Airports*” was frequently misclassified
 730 as either urban fabric, non-irrigated arable land, or pastures. These predictions likely matched the
 731 land cover of the pixel, but missed the spatial patterns that make airports easily recognizable by
 732 humans (elongated landing paths). The same issue applies to most other artificial surface LULC
 733 classes. It is also clear that the water-related classes, which are mostly distinguishable by location
 734 and spatial patterns, were often incorrectly classified as inland water bodies and maritime wetlands,
 735 which were simply more numerous in the training data.

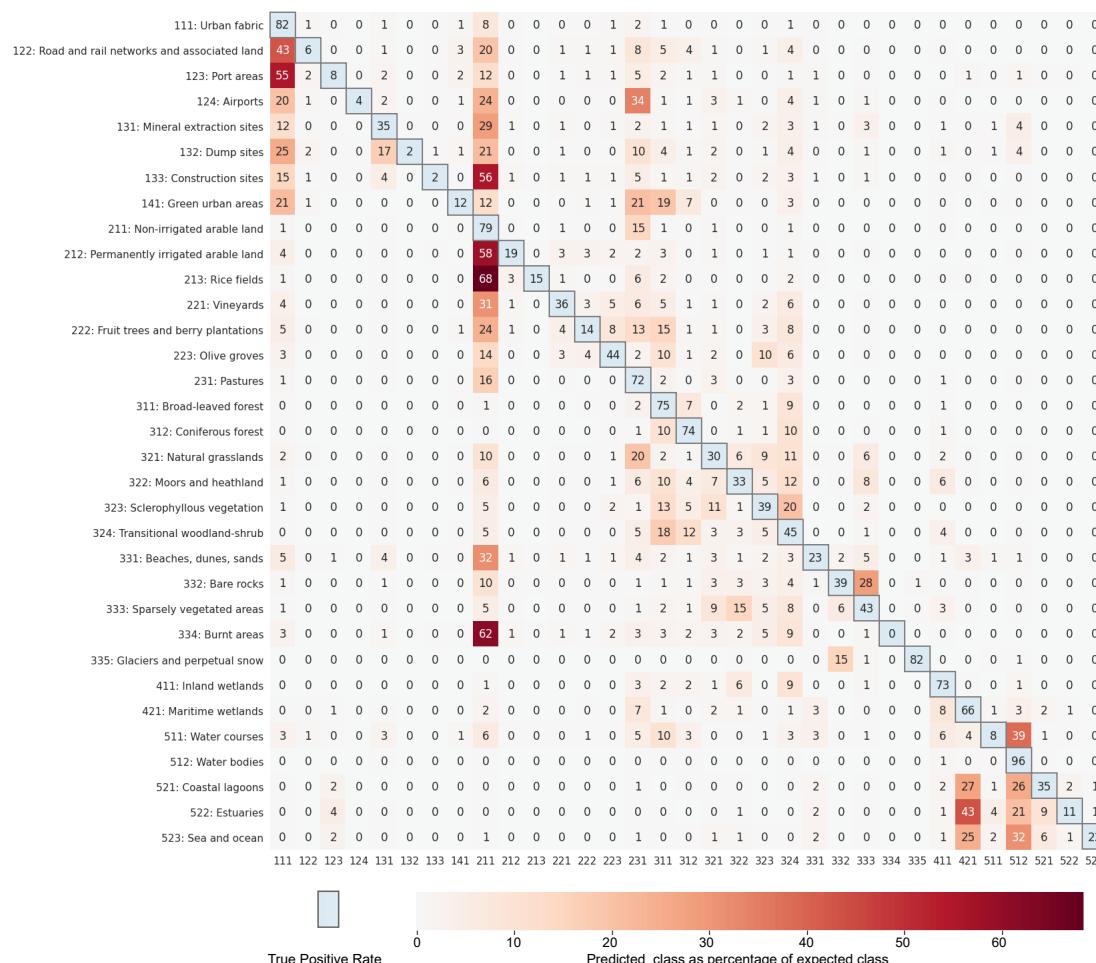


Figure 11. Percentage of predicted classes per expected class during spatial 5-fold cross-validation

736 One interesting result we noticed in this work is that training spatiotemporal models on LUCAS points

737 lead to lower classification accuracy estimates than when only using [CLC](#) points (see Table 11). This
738 was unexpected, as [LUCAS](#) land cover information stems from actual ground observations, while the
739 [CLC](#) points are pseudo-ground truth points from a dataset with a large minimum mapping unit. This
740 suggests that either the [LUCAS](#) points are harder to reproduce with remote sensing techniques, or that the
741 harmonization process needs to be improved. Further testing is needed to clarify this.

742 The field of land cover mapping is rapidly evolving. With exciting new global 10 m resolution
743 products such as [ESA](#) WorldCover and Google's Dynamic World Map expected in 2021, we expect the
744 [LULC](#) mapping bar to be raised quickly to higher resolution and higher accuracy. [Venter and Sydenham](#)
745 (2021) used low-cost infrastructure to produce land cover map of Europe at 10 m — thanks to [ESA](#) and
746 [NASA](#) making the majority of multispectral products publicly available, today everyone could potentially
747 map the world's land cover from their laptop. Our primary interest, however, will remain focusing on
748 producing better land cover maps of the past, and building land cover products that allow for unbiased
749 estimate of long-term trends. We hope that this type of data would help eventually better understand
750 the key drivers of land degradation and restoration, so that we can help stake-holders on the ground
751 make better decisions and hopefully receive financial support for the ecosystem services our environment
752 provides to us all.

753 ACKNOWLEDGEMENTS

754 OpenDataScience.eu is an Open Data project drawing inspiration from the OpenLandMap.org and
755 OpenStreetMap.org projects. This work has received funding from the European Union's the Innovation
756 and Networks Executive Agency (INEA) under Grant Agreement Connecting Europe Facility (CEF)
757 Telecom project 2018-EU-IA-0095. The authors are grateful to Radek Malinowski (Centrum Badań
758 Kosmicznych Polskiej Akademii Nauk — CBK PAN) for providing access to independent validation
759 points, and Martin Herold and Sytze de Bruin (Wageningen University) for providing suggestions on
760 methodology.

761 CONFLICT OF INTEREST

762 The authors declare no conflict of interest.

763 REFERENCES

- 764 Batista e Silva, F., Lavalle, C., and Koomen, E. (2013). A procedure to obtain a refined european land
765 use/cover map. *Journal of Land Use Science*, 8(3):255–283.
- 766 Bossard, M., Feranec, J., and Oťahel, J. (2000). Corine land cover—technical guide.
- 767 Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- 768 Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., and Smets, B. (2020). Copernicus
769 global land cover layers—collection 2. *Remote Sensing*, 12(6):1044.
- 770 Buck, O., Haub, C., Woditsch, S., Lindemann, M., Kleinwillinghöfer, L., Hazeu, G., Kosztra, B.,
771 Kleeschulte, S., Arnold, S., and Hözl, M. (2015). Analysis of the LUCAS nomenclature and proposal
772 for adaptation of the nomenclature in view of its use by the Copernicus land monitoring services.

- 773 *Machine learning*. Technical Report, European Environment Agency and the European Environment
774 Information and Observation Network.
- 775 Buus-Hinkler, J., Hansen, B. U., Tamstorf, M. P., and Pedersen, S. B. (2006). Snow-vegetation relations in
776 a high arctic ecosystem: Inter-annual variability inferred from new monitoring and modeling concepts.
777 *Remote Sensing of Environment*, 105(3):237–247.
- 778 Buyantuyev, A. and Wu, J. (2007). Effects of thematic resolution on landscape pattern analysis. *Landscape*
779 *Ecology*, 22(1):7–13.
- 780 Calderón-Loor, M., Hadjikakou, M., and Bryan, B. A. (2021). High-resolution wall-to-wall land-cover
781 mapping and land change assessment for australia from 1985 to 2015. *Remote Sensing of Environment*,
782 252:112148.
- 783 Carmel, Y., Dean, D. J., and Flather, C. H. (2001). Combining location and classification error sources for
784 estimating multi-temporal database accuracy. *Photogrammetric Engineering and Remote Sensing*, 67
785 (7): 865-872., 67(7):865–872.
- 786 Castilla, G., Larkin, K., Linke, J., and Hay, G. J. (2009). The impact of thematic resolution on the
787 patch-mosaic model of natural landscapes. *Landscape Ecology*, 24(1):15–23.
- 788 Ceccherini, G., Duveiller, G., Grassi, G., Lemoine, G., Avitabile, V., Pilli, R., and Cescatti, A. (2020).
789 Abrupt increase in harvested forest area over europe after 2015. *Nature*, 583(7814):72–77.
- 790 Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., et al. (2015).
791 Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS Journal of*
792 *Photogrammetry and Remote Sensing*, 103:7–27.
- 793 Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd*
794 *acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- 795 Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend
796 decomposition. *Journal of official statistics*, 6(1):3–73.
- 797 Conway, T. (2009). The impact of class resolution in land use change models. *Computers, Environment*
798 *and Urban Systems*, 33(4):269–277.
- 799 d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego,
800 J., Reuter, H. I., Joebges, C., Lemoine, G., et al. (2020). Harmonised lucas in-situ land cover and use
801 database for field surveys from 2006 to 2018 in the european union. *Scientific Data*, 7(1):1–15.
- 802 Defazio, A., Bach, F., and Lacoste-Julien, S. (2014). Saga: A fast incremental gradient method with
803 support for non-strongly convex composite objectives. *arXiv preprint arXiv:1407.0202*.
- 804 Duveiller, G., Caporaso, L., Abad-Viñas, R., Perugini, L., Grassi, G., Arneth, A., and Cescatti, A. (2020).
805 Local biophysical effects of land use and land cover change: towards an assessment tool for policy
806 makers. *Land Use Policy*, 91:104382.
- 807 Feng, M. and Bai, Y. (2019). A global land cover map produced through integrating multi-source datasets.
808 *Big Earth Data*, 3(3):191–219.
- 809 Feranec, J., Soukup, T., Hazeu, G., and Jaffrain, G. (2016). *European landscape dynamics: CORINE land*
810 *cover data*. CRC Press.
- 811 Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T.,
812 Daily, G. C., Gibbs, H. K., et al. (2005). Global consequences of land use. *science*, 309(5734):570–574.

- 813 Gao, B.-C. (1996). Ndwi—a normalized difference water index for remote sensing of vegetation liquid
814 water from space. *Remote sensing of environment*, 58(3):257–266.
- 815 Gao, Y., Liu, L., Zhang, X., Chen, X., Mi, J., and Xie, S. (2020). Consistency Analysis and Accuracy
816 Assessment of Three Global 30-m Land-Cover Products over the European Union using the LUCAS
817 Dataset. *Remote Sensing*, 12(21):3479.
- 818 Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts,
819 tools, and techniques to build intelligent systems*. O'Reilly Media.
- 820 Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D.,
821 Stehman, S., Goetz, S. J., Loveland, T. R., et al. (2013). High-resolution global maps of 21st-century
822 forest cover change. *science*, 342(6160):850–853.
- 823 Hengl, T., Leal Parente, L., Križan, J., and Bonannella, C. (2021). *Continental Europe Digital Terrain
824 Model at 30 m resolution based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and
825 background layers*. Zenodo.
- 826 Herold, M., Mayaux, P., Woodcock, C., Baccini, A., and Schmullius, C. (2008). Some challenges in global
827 land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote
828 Sensing of Environment*, 112(5):2538–2556.
- 829 Hillger, D., Kopp, T., Lee, T., Lindsey, D., Seaman, C., Miller, S., Solbrig, J., Kidder, S., Bachmeier,
830 S., Jasmin, T., et al. (2013). First-light imagery from suomi npp viirs. *Bulletin of the American
831 Meteorological Society*, 94(7):1019–1029.
- 832 Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel,
833 J. N., Wickham, J., et al. (2007). Completion of the 2001 national land cover database for the
834 conterminous united states. *Photogrammetric engineering and remote sensing*, 73(4):337.
- 835 Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham,
836 J., Stehman, S., et al. (2020). Conterminous united states land cover change patterns 2001–2016
837 from the 2016 national land cover database. *ISPRS Journal of Photogrammetry and Remote Sensing*,
838 162:184–199.
- 839 Hong, C., Burney, J. A., Pongratz, J., Nabel, J. E., Mueller, N. D., Jackson, R. B., and Davis, S. J. (2021).
840 Global and regional drivers of land-use emissions in 1961–2017. *Nature*, 589(7843):554–561.
- 841 Houghton, R. A., House, J. I., Pongratz, J., Van Der Werf, G. R., DeFries, R. S., Hansen, M. C.,
842 Quéré, C. L., and Ramankutty, N. (2012). Carbon emissions from land use and land-cover change.
843 *Biogeosciences*, 9(12):5125–5142.
- 844 Huang, K., Xia, J., Wang, Y., Ahlström, A., Chen, J., Cook, R. B., Cui, E., Fang, Y., Fisher, J. B.,
845 Huntzinger, D. N., et al. (2018). Enhanced peak growth of global vegetation and its key mechanisms.
846 *Nature ecology & evolution*, 2(12):1897–1905.
- 847 Huete, A. R. (1988). A soil-adjusted vegetation index (savi). *Remote sensing of environment*, 25(3):295–
848 309.
- 849 Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational high resolution
850 land cover map production at the country scale using satellite image time series. *Remote Sensing*,
851 9(1):95.
- 852 Jin, S. and Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized

- 853 difference moisture index in detecting forest disturbances. *Remote sensing of Environment*, 94(3):364–
854 372.
- 855 Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., and Goldewijk, K. K.
856 (2011). Holocene carbon emissions as a result of anthropogenic land cover change. *The Holocene*,
857 21(5):775–791.
- 858 Key, C. H. and Benson, N. C. (1999). Measuring and remote sensing of burn severity. In *Proceedings
859 joint fire science conference and workshop*, volume 2, page 284. University of Idaho and International
860 Association of Wildland Fire Moscow, ID.
- 861 Key, C. H. and Benson, N. C. (2006). Landscape assessment (la). In: Lutes, Duncan C.; Keane, Robert E.;
862 Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006. FIREMON:
863 *Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-GTR-164-CD*. Fort Collins, CO:
864 US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. LA-1-55, 164.
- 865 Kilibarda, M., Hengl, T., Heuvelink, G. B., Gräler, B., Pebesma, E., Perčec Tadić, M., and Bajat, B.
866 (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution.
867 *Journal of Geophysical Research: Atmospheres*, 119(5):2294–2313.
- 868 Korhonen, K. T. (2020). *A new article in the journal Nature overestimates the increase of forest harvesting
869 in Europe*. Natural Resources Institute Finland (Luke). Accessed: 2021-07-04.
- 870 Li, X., Zhou, Y., Zhu, Z., Liang, L., Yu, B., and Cao, W. (2018). Mapping annual urban dynamics
871 (1985–2015) using time series of landsat data. *Remote Sensing of Environment*, 216:674–683.
- 872 Liu, H., Gong, P., Wang, J., Clinton, N., Bai, Y., and Liang, S. (2020a). Annual dynamics of global land
873 cover and its long-term changes from 1982 to 2015. *Earth System Science Data*, 12(2):1217–1243.
- 874 Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J. (2021). Finer-resolution mapping of global
875 land cover: Recent developments, consistency analysis, and prospects. *Journal of Remote Sensing*,
876 2021.
- 877 Liu, X., Trogisch, S., He, J.-S., Niklaus, P. A., Bruelheide, H., Tang, Z., Erfmeier, A., Scherer-Lorenzen,
878 M., Pietsch, K. A., Yang, B., et al. (2018). Tree species richness increases ecosystem carbon storage in
879 subtropical forests. *Proceedings of the Royal Society B*, 285(1885):20181240.
- 880 Liu, Y., Hou, X., Li, X., Song, B., and Wang, C. (2020b). Assessing and predicting changes in ecosystem
881 service values based on land use/cover change in the bohai rim coastal zone. *Ecological Indicators*,
882 111:106004.
- 883 Lovelace, R., Nowosad, J., and Muenchow, J. (2019). *Geocomputation with R*. Chapman & Hall/CRC
884 The R Series. CRC Press.
- 885 Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski,
886 A., Wojtkowski, C., Krupiński, M., Krätzschmar, E., et al. (2020). Automated Production of a Land
887 Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sensing*, 12(21):3523.
- 888 McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The
889 bulletin of mathematical biophysics*, 5(4):115–133.
- 890 Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good
891 practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*,
892 148:42–57.

- 893 Palahí, M., Valbuena, R., Senf, C., Acil, N., Pugh, T. A., Sadler, J., Seidl, R., Potapov, P., Gardiner,
894 B., Hetemäki, L., et al. (2021). Concerns about reported harvests in european forests. *Nature*,
895 592(7856):E15–E17.
- 896 Paulsson, J., Claesson, S., Fridman, J., and Olsson, H. (2020). Incorrect figures on harvested forests in
897 nature article. *SLU news*. Accessed: 2021-07-04.
- 898 Payn, T., Carnus, J.-M., Freer-Smith, P., Kimberley, M., Kollert, W., Liu, S., Orazio, C., Rodriguez, L.,
899 Silva, L. N., and Wingfield, M. J. (2015). Changes in planted forests and future global implications.
900 *Forest Ecology and Management*, 352:57–67.
- 901 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer,
902 P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of
903 machine Learning research*, 12:2825–2830.
- 904 Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global
905 surface water and its long-term changes. *Nature*, 540:418–.
- 906 Pflugmacher, D., Rabe, A., Peters, M., and Hostert, P. (2019). Mapping pan-European land cover using
907 Landsat spectral-temporal metrics and the European LUCAS survey. *Remote sensing of environment*,
908 221:583–595.
- 909 Pielke Sr, R. A., Marland, G., Betts, R. A., Chase, T. N., Eastman, J. L., Niles, J. O., Niyogi, D. D. S., and
910 Running, S. W. (2002). The influence of land-use change and landscape dynamics on the climate system:
911 relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical
912 Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering
913 Sciences*, 360(1797):1705–1719.
- 914 Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B.,
915 Tyukavina, A., and Ying, Q. (2020). Landsat analysis ready data for global land cover and land cover
916 change mapping. *Remote Sensing*, 12(3):426.
- 917 Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted
918 vegetation index. *Remote sensing of environment*, 48(2):119–126.
- 919 Raši, R. (2020). *State of Europe's Forests 2020*. Ministerial conference on the protection of forests in
920 Europe. Accessed: 2021-10-05.
- 921 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-
922 Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with
923 temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.
- 924 Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes,
925 E. C., Pandey, B., et al. (2018). Nasa's black marble nighttime lights product suite. *Remote Sensing of
926 Environment*, 210:113–143.
- 927 Rossi, F., Breidenbach, J., Puliti, S., Astrup, R., and Talbot, B. (2019). Assessing harvested sites in a
928 forested boreal mountain catchment through global forest watch. *Remote Sensing*, 11(5).
- 929 Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E.,
930 Huenneke, L. F., Jackson, R. B., Kinzig, A., et al. (2000). Global biodiversity scenarios for the year
931 2100. *science*, 287(5459):1770–1774.
- 932 Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In

- 933 9th Python in Science Conference.
- 934 Senf, C., Pflugmacher, D., Zhiqiang, Y., Sebald, J., Knorn, J., Neumann, M., Hostert, P., and Seidl, R.
935 (2018). Canopy mortality has doubled in europe's temperate forests over the last three decades. *Nature
936 Communications*, 9(1):1–8.
- 937 Senf, C. and Seidl, R. (2021). Mapping the forest disturbance regimes of europe. *Nature Sustainability*,
938 4(1):63–70.
- 939 Seni, G. and Elder, J. (2010). *Ensemble Methods in Data Mining: Improving Accuracy Through
940 Combining Predictions*. Synthesis lectures on data mining and knowledge discovery. Morgan &
941 Claypool Publishers.
- 942 Shumba, T., De Vos, A., Biggs, R., Esler, K. J., Ament, J. M., and Clements, H. S. (2020). Effectiveness
943 of private land conservation areas in maintaining natural land cover and biodiversity intactness. *Global
944 Ecology and Conservation*, 22:e00935.
- 945 Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., and Townshend,
946 J. R. (2018). Global land change from 1982 to 2016. *Nature*, 560(7720):639–643.
- 947 Song, X.-P., Sexton, J. O., Huang, C., Channan, S., and Townshend, J. R. (2016). Characterizing
948 the magnitude, timing and duration of urban growth from time series of landsat-based estimates of
949 impervious cover. *Remote Sensing of Environment*, 175:1–13.
- 950 Sy, S. and Quesada, B. (2020). Anthropogenic land cover change impact on climate extremes during the
951 21st century. *Environmental Research Letters*, 15(3):034002.
- 952 Szantoi, Z., Geller, G. N., Tsedbazar, N.-E., See, L., Griffiths, P., Fritz, S., Gong, P., Herold, M., Mora,
953 B., and Obregón, A. (2020). Addressing the need for improved land cover map products for policy
954 support. *Environmental Science & Policy*, 112:28–35.
- 955 Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M.,
956 Narasimhan, R., Kim, D., et al. (2012). Global characterization and monitoring of forest cover using
957 landsat data: opportunities and challenges. *International Journal of Digital Earth*, 5(5):373–397.
- 958 Trisurat, Y., Shirakawa, H., and Johnston, J. M. (2019). Land-use/land-cover change from socio-economic
959 drivers and their impact on biodiversity in nan province, thailand. *Sustainability*, 11(3):649.
- 960 Tsedbazar, N., Herold, M., De Bruin, S., Lesiv, M., Fritz, S., Van De Kerchove, R., Buchhorn, M.,
961 Duerauer, M., Szantoi, Z., and Pekel, J.-F. (2018). Developing and applying a multi-purpose land cover
962 validation dataset for africa. *Remote Sensing of Environment*, 219:298–309.
- 963 Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation.
964 *Remote sensing of Environment*, 8(2):127–150.
- 965 Van, T. T., Cao, D. P., Nishida, N. K., et al. (2019). How does land use/land cover map's accuracy depend
966 on number of classification classes? *SOLA*, 15:28–31.
- 967 Van Rijsbergen, C. (1980). *Information Retrieval*. Butterworth Heinemann.
- 968 Veldkamp, A. and Lambin, E. F. (2001). Predicting land-use change. *Agriculture Ecosystems and
969 Environment*.
- 970 Venter, Z. S. and Sydenham, M. A. K. (2021). Continental-scale land cover mapping at 10 m resolution
971 over europe (elc10).
- 972 Vilar, L., Garrido, J., Echavarria, P., Martinez-Vega, J., and Martin, M. P. (2019). Comparative analysis of

- 973 CORINE and climate change initiative land cover maps in Europe: Implications for wildfire occurrence
974 estimation at regional and local scales. *International Journal of Applied Earth Observation and*
975 *Geoinformation*, 78:102–117.
- 976 Wang, X., Wu, C., Peng, D., Gonsamo, A., and Liu, Z. (2018). Snow cover phenology affects alpine
977 vegetation growth dynamics on the tibetan plateau: Satellite observed evidence, impacts of different
978 biomes, and climate drivers. *Agricultural and Forest Meteorology*, 256:61–74.
- 979 Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Springer New
980 York.
- 981 Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J. (2020). GLC_FCS30: Global land-cover product
982 with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data*
983 *Discussions*, pages 1–31.
- 984 Zhou, W., Qian, Y., Li, X., Li, W., and Han, L. (2014). Relationships between land cover and the surface
985 urban heat island: seasonal variability and effects of spatial and thematic resolution of land cover data
986 on predicting land surface temperatures. *Landscape ecology*, 29(1):153–167.
- 987 Zhu, Z. and Woodcock, C. E. (2014). Continuous change detection and classification of land cover using
988 all available landsat data. *Remote sensing of Environment*, 144:152–171.